DOCUMENT RESUME

ED 192 168 CE 027 163

AUTHOR	Huff, Sheila M.: And Others
TITLE	The Assessment of Occupational Competence. 4. The Definition and Measurement of Competence in Higher Education.
INSTITUTION	McBer and Co., Boston Mass.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Feb 80
CONTRACT	400-78-0028
NOTE	170p.: For related documents see CE 027 159-165.
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Academic Achievement: Comparative Analysis: *Competence: *Competency Based Education: *Demonstration Programs: *Educational Testing: Education Work Relationship: Higher Education: Job Skills: Minimum Competency Testing: Postsecondary Education: Program Descriptions: Secondary Education
IDENTIFIERS	*Occupational Competence Assessment

ABSTRACT

Focusing on the competency based education movement (CBE), this chapter is one of seven sections of a report that examines the assessment of occupational competence. The chapter begins by reviewing the prominent CBE programs, their institutional objectives, and how competencies are identified and assessed. In particular, the degree to which institutional goals influence the identification and assessment process is examined, with traditional programs as a comparison. Some of the questions to which the chapter responds are the following: What are the distinctions between CBE and more traditional programs in the way they go about assessing competence? How does the specification of competencies shape curriculum content? How does the practice of assessment affect student learning? And how do assessment procedures limit or distort the competencies that are defined as curriculum objectives? Following the review of the state of the practice, a model for competency definition is introduced. The model becomes the framework for a discussion of the practical directions that institutions of higher education can take in pursuing their goals, within some of the real-world limitations imposed by tradition, the faculty, time, and financial resources. (Other sections of the report are available separately--see note. The first is an overview: the last is a synthesis of issues.) (LRA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192168

CE027 163

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE ASSESSMENT OF OCCUPATIONAL COMPETENCE

4. THE DEFINITION AND MEASUREMENT OF COMPETENCE IN HIGHER EDUCATION

Sheila M. Huff
George O. Klemp, Jr.
David G. Winter

Prepared for
The National Institute of Education
Contract NIE 400-78-0028

February, 1980

McBER 137 NEWBURY STREET · BOSTON · MASSACHUSETTS 02116

DEC 4 1980

TABLE OF CONTENTS

page

I. Introduction 4.1

II. Brief History of the CBE Movement 4.6

1. Competency-Based Teacher Education
2. Competency-Based Education (CBE)
3. CBE at the Secondary Level
4. Competency-Based Vocational Education
5. Credit for Experiential Learning

III. Assessing the State of Practice 4.26

- Alverno College
- College for Human Service
- Kirkhof College
- Florida State University
- Mt. Hood Community College, Gresham, Oregon
- Antioch School of Law

IV. The Sources and Nature of Competence Goals 4.56

- Harvard College
- Alverno College
- Kirkhof College
- Southern Illinois University School of Medicine
- Antioch School of Law
- College for Human Service
- The Practice of Assessing Competency Attainment
- The Nature of the Competencies Measured
- The Explicitness of the Assessment Criteria

V. An Alternative Approach to CBE 4.105

- Steps in the Process
- How the Process Relates to Assessment in Current CBE Programs
- Validating Measures of Program Effectivenss
- Some New Measures of Educational Effects

VI. Conclusions 4.147

I. INTRODUCTION

The broad context of this chapter is national interest in the relationship between education and work. That relationship has myriad dimensions which generate weighty policy issues. One concern, which has occupied a great deal of federal and state legislators of late, is what to do to assure that students have "saleable skills." Momentum behind legislative interest in this area comes primarily from worried parents and students: "Will there be enough jobs to go around?" "What should be learned to be in a good competitive position in the labor market?"

The need for a closer look at "outcome-oriented" education springs from a sense of urgency developing in Congress about education and work problems, particularly youth unemployment. Also, competency-based education (CBE) was beginning to be promoted on a large scale as an answer to the "saleable skills" question, or as a broad-spectrum antibiotic capable of curing many perceived ailments of the educational system. Assumptions were and are being made about the model's potential; a major one is that employers and certifying authorities know what is needed for competency and all that must happen is to find ways for them to educate school system personnel.

The chapters on employment selection, assessment centers, and licensing and certification practices demonstrate the falseness of that assumption. These systems do not operate on knowledge of what is important to measure and how to measure it. Rather, they operate for the most part on beliefs (or assumptions) about what is important to measure and how to measure it: some beliefs undoubtedly well-founded, others are not. Moreover, activity in the courts had been expected to promote growth in the knowledge base: Instead, developments seem to be leading employers in the opposite direction. Given these realities, in what ways have CBE program developers defined and measured competence? Can employment selection and training profit from the work that has been done in the education sector, or *vice versa*? Where are the gaps, and what can be done to improve the situation?

We begin our examination of CBE by reviewing the relatively short history of the movement, tracing its emergence in teacher education, undergraduate college programs, general education programs in secondary schools, and vocational and technical education at the secondary level. For a further discussion of the philosophical roots of the movement, the reader is referred to Grant, et al. (1979).

Throughout the following section of this chapter, the focus is on competency-based postsecondary education, due to both the keen interest of policymakers in this area and the fact that most of what may be considered central to the "state of practice" is to be found here. Our review touches on 18 individual CBE programs, a representative sample of which are described in detail. In each of these programs, three themes emerge which may be taken as espoused ideals of CBE programs:

1. The goals of education should be conceptualized as effective actions or performance in some later role or milieu, rather than inferred cognitive or affective internal states such as "knowledge" (either of specific facts or of more abstract general principles), or "maturity."

2. Educational goals should be set forth in language that is as explicit as possible, broken down as far as possible into component outcomes which lend themselves to measurement.

3. Student achievement should be assessed in terms of the ability to demonstrate "the behavior itself", that is, performances that are more or less isomorphic to performances in the later milieu or roles for which students are being educated.

When the CBE movement is measured against these ideals, it will be seen that it is a set of interlocking emphases, rather than sharply defined separate entities. The three essential features described above are really the ends of continua along which a variety of educational programs can be arrayed. There is also nothing in particular that makes these ideals the exclusive province of the CBE movement. For example, almost all Ph.D. programs and many undergraduate honors programs in traditional liberal arts institutions require a demonstration of "scholarly performance" through research and a thesis that is debatably as isomorphic to "real world" scholarship and research as are the performances in any CBE program.

As many of the essential features of CBE might be expected to be found in other kinds of higher education, it was necessary to frame broader questions and cast a wider net, setting CBE within the context of higher education as a whole. The comparison of CBE with traditional undergraduate programs frames the conclusions regarding the state of practice in CBE.

Ideals being ideals, there are numerous gaps between where CBE is presently and where it should be: The movement, after all, has only been in existence a short time and has yet to realize its potential. Based on our synthesis of the current state of practice, and referring to alternative models for CBE in use outside of college programs, the third section of this chapter presents a model of curriculum and assessment development based on empirically demonstrated notions of competence. Though the organizations that have implemented or are implementing variations on this model are focusing on perhaps a narrower range of goals than those espoused by the college programs surveyed (including management, leadership, and technical skills excellence), the model may serve as a useful point of comparison for understanding how CBE can be defined and implemented in different settings. This model will then be applied to current practices in CBE to examine: (1) what benefits, if any, might be derived from a systematic, empirically-based strategy of competency identification; and (2) what constraints of practicality and institutional mission suggest as workable, realistic advances in the state of practice. We add a discussion of some empirically-derived measures of competencies which have been shown in research to have demonstrated relevance to the ideals of CBE.

Our analysis is based on published reports, handbooks, and other materials from the institutions themselves; conclusions and reports prepared by other observers; visits to campuses; and a wide range of personal contacts with people within CBE. This report is certainly not an exhaustive compilation of the goals and assessment procedures of every conceivable CBE program, nor is it intended as such. It is rather an analytic delineation of the most important dimensions along which all educational goals and procedures can be described, highlighted by a comparison of representative institutions along these dimensions. From such an analysis we believe it is possible to draw significant conclusions about the definition and measurement of "competence" in American higher education today.

II. BRIEF HISTORY OF THE CBE MOVEMENT

The word "competence" was borrowed from ordinary language and transformed into a technical term by psychologists, beginning with the seminal article, "The Concept of Competence," by R. W. White (1959). "Competence" has since become a public slogan reflecting concerns for improved quality control in instruction and selection systems. What follows is a look at five tracks which have emerged over the last ten years to address competence goals in education: (1) competency-based teacher education; (2) CBE in other postsecondary institutions; (3) CBE at the secondary level; (4) vocational and technical school approaches to CBE; and (5) credit for experiential learning. This multifaceted look at the history of CBE develops the context for our study of the current practice of competence assessment in the educational arena.

1. Competency-Based Teacher Education*

CBE was first developed as a conceptual framework or model for programs in the field of teaching. In the early 1970s, the accountability movement in education was in full swing.

---

*The terms "competency-based" and "performance-based" are often used interchangeably. "Performance-based teacher education" (PBTE) was used most frequently in the early 1970s, while "competency-based teacher education" (CBTE) is most current.

Legislatures across the country were mandating statewide testing of elementary and secondary students and were demanding that schools demonstrate their effectiveness in terms of student learning outcomes, especially in basic skills areas (e.g., literacy). To establish the targets of statewide testing, goals for elementary and secondary schools were re-examined. Panels of parents, educators, school personnel and government officials were assembled in many states to address questions, such as what children should be taught and what counts as a basic skill (Clasby, et al., 1973). In New Jersey, it was mandated that all children in the state receive a "thorough and efficient education," although the meaning of "thorough and efficient education" is still being debated there.

In waves of increasing criticism of the state of educational practice, reflected in such book titles as Why Johnny Can't Read and Crisis in the Classroom, a common charge was that faculty of schools of education spent little if any time in primary or secondary schools and did not know much about life in classrooms; therefore, what they taught their students was to be treated as suspect. Critics proposed that teacher education be firmly grounded in knowledge of the demands of practice and prerequisites of effective performance, thus giving rise to the competency-based teacher education (CBTE) movement.

State education departments and legislatures across the country began exploring, advocating or mandating CBTE (Schmieder, 1974). Also, the then prevalent form of teacher certification--the program approval method--was to be reconsidered.* Under that system, anyone who completed an accredited teacher training program, which usually included a supervised field placement component, was automatically certified by the state. This kind of certification system is inexpensive to administer. A desire to maintain the program approval system, served as an incentive to mandating CBTE. It was proposed that the program approval system was justifiable only if competent novices were graduated by schools of education.

The system surrounding the education of teachers, then, was the first to explore CBE. Hopes ran high in the early years that with reasonable investments in research on teacher effectiveness, CBTE could become the prevalent model by the end of the decade. New York state, a leader in the CBTE movement, created the goal of 1980 for all teacher training in the state to be competency-based and for those programs to be supported by a firm research base (Schmieder, 1974; Huff, 1976).

Although there was early optimism, leaders of the CBTE movement rapidly discovered the realities of the situation. In 1974, in a publication that was part of a series on PBTE, (Merwin, 1973), David Krathwohl wrote:

---

*Some states and locales used different systems such as certification and/or licensing contingent upon acceptable examination scores or acceptable performance during probationary periods of employment.

> One can predict that performance-based teacher education (PBTE) is certain to fail to reach its ultimate objective if it continues on its present course. This failure will be caused by the almost complete lack of attention given to the assessment of teaching competencies, a core concept of PBTE.

The authors of a 1975 publication in the same series comment on Krathwohl's diagnosis:

> We suspect that this neglect of the assessment problems is not entirely the result of conflicting demands on program developers' time. Our contacts with these harried individuals reveal that they have a strong reluctance to tackle these problems because of feelings of inadequacy arising from the lack of knowledge of adequate approaches to solutions. (Medley, et al., 1975)

Under the pressure of mandates and state promotional campaigns, a common response was to assemble panels to define teacher competencies. The results were usually taxonomies of hundreds of tasks and behaviors that teachers supposedly had to be able to perform. Ralph Tyler, a pioneer in developing behavioral objectives for learning, noted that these products were regressions to the Commonwealth Foundation of New York's Teacher Training Study of 1928, which listed 1001 activities of teachers based on reports from teachers across the state on areas they considered important to proficiency. About this and current lists, Tyler (1980) said they will always be trivial until and unless analysis proceeds with a causal understanding of what enables people to perform the tasks competently: what enables a teacher, for example, to modify his or her behavior in light of new information and in new situations. He noted that most implementors of ideas such as "behavioral objectives" and CBTE are process-oriented and lack a theoretical understanding of teaching and learning. Under these circumstances, the results can be expected to be atomistic, fragmentary and not trustworthy.

To underscore this point, we note that a special "legal issues" conference was sponsored in 1976 by the U.S. Office of Education for the Multi-State Consortium on Performance-Based Teacher Education. State education personnel, members of the research and development community, teacher union representatives, and members of PBTE task forces and study groups met to, among other things, vent frustration at the grossly deficient amount of solid research on teacher competencies and to worry together about such things as whether CBTE institutions might be sued if their graduates were incompetent (Huff, 1976). One institution asked if it could, in this event, lodge a countersuit against the state, since "the state made us do it." The institution subsequently defined itself as "competence-oriented," and the state education department is keeping quiet on the topic of CBTE.

There have been several recent critical reviews of teacher effectiveness research which pinpoint methodological problems and testify to the lack of useful information to guide program development for pre-service or in-service teacher education (Borich, 1977; Shavelson & Atwood-Russo, 1977; Shavelson & Dempsey, 1976; Winne & Marx, 1977). Moreover, because of rising public concern over basic skills education (the "Three R's"), the student outcomes that have been measured rigorously in relationship to teacher characteristics or behaviors have been primarily cognitive, associated with skills in reading and mathematics (e.g., the Beginning Teacher Evaluation Study directed by Frederick J. McDonald of ETS (1976), sponsored by the National Institute of Education's Teaching and Learning Division). For this reason, Borich (1977) found it necessary to remind the research and development community of the importance of affective as well as cognitive student outcomes, and to point out that positive effects in the cognitive domain may have either no effect or a negative effect in the affective domain (and vice-versa). Borich went on to say that while researchers and evaluators recognize these points, most fail to adapt their instrumentation to the full range of outcome behavior that could be measured. Research questions, rather than driving the scope of methods of inquiry, are defined and constrained by customarily used or available methods. This phenomenon is commonly referred to as "the tail wagging the dog." Competency-based teacher education is still attempting to overcome it. Few programs have reduced the effect to any significant extent.*

---

*Personal communications with New York State Education Department personnel and multi-state consortium on PBTE members led to this general conclusion.

Despite rude awakenings and big disappointments, CBTE has survived. The lack of a research base is now generally acknowledged, but the model is still being promoted on pragmatic grounds, as it was in 1976 (Huff, 1976). It encourages faculty of schools of education to learn more about classrooms and problems that teachers confront. It is hoped that faculty will spend more time in schools listening to teachers, students, and parents and that this, in turn, will yield curricula that are performance-based, at least, if not wholly grounded on knowledge of what is needed for competence.

It should be noted, however, that a counter-trend to CBTE is developing in some states in the wake of the United States v. South Carolina case.* Out of concern for institutional accreditation and the labor market opportunities of their graduates, schools of education are retreating, to the dismay of some education officials, from an interest in or orientation to performance and are centering their curricula on objectives that assure passing scores of their students on the National Teachers Examination (NTE). For the South Carolina case, the courts accepted as a demonstration of content validity for use of the measure in certifying teachers, ETS's evidence that the NTE was related to the curricula in South Carolina institutions.**

---

*445 F. Supp. 1094 (D.S.C. 1977), aff'd, 434 U.S. 1026 (1978) See Chapter 5, section on CBE and the NTE.

**445 F. Supp. at 1112.

2. Competency-Based Education (CBE)

Another independent strand of the competency-based education movement can be traced to a special focus program on competency-based education of the Fund for the Improvement of Postsecondary Education (FIPSE), an agency of the Department of Health, Education, and Welfare.

FIPSE was a new agency established in recognition of several decades of relative inattention of federal agencies to higher education. FIPSE's staff, under the early leadership of Virginia Smith and Russell Edgerton, wished to fund reforms but did not wish to encourage fads to be passed from campus to campus. The idea of "outcome driven," "student-centered" programs was attractive, since institutions would closely examine their goals and missions. Programs would then flow from concrete statements of purpose.

In 1972, FIPSE cast its net in search of proposals that showed promise for meaningful, enduring reform. Two submissions created considerable excitement among agency staff, since they defined "outcome driven" and "student-centered" in concrete terms. These were the proposals of Alverno College and College III, University of Massachusetts. Both institutions used the terminology of "competence." For the following year, guidelines were developed at FIPSE for a special focus program on competency-based education.

Having funded several proposals to develop CBE programs, FIPSE sponsored the Task Force on CBE in 1973. The Task Force sought to better understand the nature of innovations under the banner of CBE and to suggest ways that the research and development community could provide assistance to CBE practitioners.

Until this time, FIPSE was totally unaware of competency based teacher education. The only significant crossfertilization was through Frederick J. McDonald, who served on FIPSE's Task Force. McDonald, a social scientist who was conducting a major study on teacher effectiveness and directing a national commission on PBTE, counseled the Task Force on basic research issues. The Task Force concluded early on that, as with CBTE, the research base for CBE programs was practically non-existent. CBE had to be temporarily justified on other pragmatic grounds, as did CBTE. In many of the FIPSE projects, new students were being served, institutional missions and goals were carefully re-examined, serious thought and considerable energy were being devoted to pedagogy, and the tail (i.e., available assessment measures) was not wagging the dog.

CBE practitioners were, for the most part, fending for themselves in developing definitions of competence and means of assessment. The second section of this chapter reports on subsequent developments in competency-based postsecondary programs over the past seven years.

3. CBE at the Secondary Level

There are several factions of the competency-based education movement at the secondary level. "Minimum competency testing"* is well known to the public as the most current manifestation of the accountability movement. As of the present, according to data from the Education Commission of the States in Denver, 38 states have some kind of minimum competency testing mandated by state boards of education or state legislatures. Eighteen of the 38 states require passing tests to receive a high school diploma.

In the case of minimum competency tests, the "competency" label is somewhat misleading, since most of the tests in use are standardized paper-and-pencil achievement tests which tap a narrow range of cognitive ability in communications and computation. However, in reference to competence, justification for the content of the tests is sometimes given in terms of the University of Texas's Adult Performance Level (APL) research, a five-year program which assessed the competencies of the general population (Nortcutt, et al., 1975). Together with news from the College Entrance Examination Board of declining SAT scores (Ryan, 1977) and a combination of other influences, the APL study's statistics provided impetus for state and local minimum competency testing programs. The APL report, based on a national sample of 1500 adults aged 18 to 65, examined five content areas (consumer economics, occupational knowledge, community resources, health, and government and law) and four skill areas (communications, computation, problem solving, and interpersonal skills). Among other things, the study released these shocking estimates:

---

*See Chapter 5 by for a discussion of legal issues surrounding minimum competency testing, major issues being instructional validity of the tests and due process in phasing in minimum competency testing programs. For a discussion of pros and cons of minimum competency testing, see Irby 1979.

- 52 million adults cannot correctly match personal qualifications to job requirements in classified advertisements.
- 48 million adults cannot determine the correct amount of change from a purchase when given a cash register receipt and the denomination of the money to be used.
- 39 million adults cannot interpret an earnings statement well enough to find the deduction for social security.
- 23 million adults lack important functional competencies, and at least 34 million more are functioning but not proficient in the five content and four skill areas of competence identified in the study.

Concurrent with the APL research study, the New York State External High School Diploma Program was conceived and implemented. In 1972, a task force of college, business, industry, and education community representatives, headed by Stephen K. Bailey, Francis U. Macy, and Donn F. Vickers of Syracuse Research Corporation's Regional Learning Service (SRC/RLS), defined necessary adult competency areas. The competency areas are almost identical to those that emerged later from the APL research.

The New York State External Diploma Program, administered through SRC/RLS, emphasizes assessment or ratification of adult competencies. Unlike most competency-based external diploma programs, the New York State program has no instructional component. Instead, a comprehensive local educational resource directory is used in counseling adults on where they might seek assistance in improving their levels of proficiency (this process is known as "educational brokering"). Central to the program's conception was the idea of providing alternate routes to credentials needed for access to jobs and promotions. According to Nickse and McClure (1980), when programs have an instructional component, there is usually little, if any, investment in developing performance-based assessments. In contrast, the assessments of the New York State program are all performance-based. Operational definitions of competence and assessment measures were developed by the first and second assessment directors of the program, Ruth S. Nickse and Judith Alamprese, respectively (Alamprese, 1979; Nickse, 1980). As originally conceived, the program was to have been the first regional certification that could serve as a model for others at the secondary and postsecondary levels.

The Texas APL project was among the first programs to borrow heavily from the operational definitions and assessments developed for New York State. With training provided by SRC/RLS staff, the New York State model has since been implemented in Wisconsin, Maryland, Connecticut, and Michigan.

Training is soon to begin in Montana, and education department personnel in several other states are studying the model's applicability to their needs.

The New York State model and the Texas APL model have recently been approved for national dissemination by the Joint Dissemination Review Panel of the U.S. Office of Education and National Institute of Education. These are the only two competency-based high school programs approved for the National Diffusion Network. While both are oriented to adults, the models provide valuable information for youth programs as well.

Most nonvocational (general education) competency-based secondary programs have borrowed only that aspect of CBE which calls for explicitness about teaching objectives. This has been done primarily to satisfy the accountability needs of districts and states. Carnegie units and state-mandated subject area requirements are still the mainstays of the system, and almost all of these programs use traditional paper-and-pencil, teacher-constructed examinations. Also, the schools' general education components must serve students who are college bound as well as those directly entering jobs after high school. This complicates issues, since the competencies needed for success in higher education programs are not assumed to be the same as those needed for success in securing employment and profiting from on-the-job training opportunities. A parallel program for the two groups, with early tracking of students, is simply not thought to be a feasible solution to this dilemma.

These are but a few of the problems and constraints which serve to limit expectations for further developments in competency-based education for comprehensive high school programs.

4. Competency-based Vocational Education

Competency-based vocational education, on the other hand, goes far beyond specificity and explicitness about educational objectives, historically the norm for such programs. Most of these programs are constructing or rebuilding their curricula on task analyses of jobs. A major resource for task analyses is the Vocational/Technical Education Consortium of States (V-Tecs), based in Atlanta, Georgia. The Consortium was founded in July of 1973 with an initial membership of seven states. Presently, 16 states are members, and the U.S. Army, Navy, and Air Force hold associate memberships. V-Tecs' purpose is to provide sound information from task analyses to program planners and curriculum developers. The membership fee is $20,000 per year. Membership entitles a state to the V-Tecs catalogs which contain performance objectives and guides in the selected domain or area, and the new member must, in turn, produce a catalog in an area judged to be a high priority by V-Tecs.

V-Tecs catalogs are the end result of task analyses of specific job titles, coded for access in accordance with the U.S. Department of Labor's Dictionary of Occupational Titles and the U.S. Office of Education's vocational coding system. Sixty-seven catalogs are available, and an additional 59 are currently being produced. For example, if the task is to take a blood pressure count, the performance objective might read: "Given a blood cuff and a patient, the student/nurse will measure the patients's blood pressure to an accuracy of ___X___." The performance guide would then go on to list all the component steps in executing the task. Tools or equipment used to perform the task are also listed.

Not all identified tasks are selected for presentation in this manner. Judgments as to which merit thorough attention are made by state education department staff. The catalog, however, includes the full list of identified tasks together with statistical information from the surveys and references from the literature reviews conducted.

Steps in developing the catalogs are as follows:

1. State technical staff (based upon the state's formal plan for vocational education) and V-Tecs decide on a "domain/area" that meets their needs.
2. A literature review is conducted for the "domain/area."
3. The population to be sampled is identified (i.e., names and addresses of journeyman electricians are obtained from unions).
4. State vocational education coordinators and staff conduct interviews of a small sample of workers to identify tasks, tools and equipment.
5. Using information from the literature review and interviews, state staff develop a survey instrument.
6. A random sample of workers is asked if they perform the tasks. The workers are also asked to rate the average duration of each task and to verify the tools and equipment lists.

7. Working from the survey data and literature review, the state writing team selects which tasks will be converted for the catalog, writes the performance objectives and guides for the selected tasks, and produces a draft of the catalog which is reviewed by a sample of workers and revised.

Program developers are urged by V-Tecs to verify the appropriateness of the tasks for their locales and goals. The work of putting the tasks in sequence and organizing learning experiences around them is left to curriculum developers at the local level, as is the work of infering the knowledge and attitudes necessary to demonstrate the skills evident in the tasks (Rohrback, 1980).

Needless to say, not all developing competency-based vocational programs use the V-Tecs materials. A prevalent mode for developing performance objectives, especially where V-Tecs materials are not available, is a panel approach to task analysis. Panel members may be selected with care to include workers themselves and to ascertain whether the workers know their jobs well and are successful in performing them; or panels may be hastily thrown together in an ad hoc fashion. Also, in terms of designing survey instruments or simple task lists for review by workers or interview protocols for gathering data, program developers usually lack appropriate training. And of course, most programs do not have sufficient funds to conduct surveys.

"Borrowing" of materials and ideas--wholesale or with adaptation--by harried program developers often occurs to fill voids. The competency-based vocational programs in suburban Hennepin and White Bear Lake, Minnesota, have provided valuable materials and guidance on all aspects of program development and implementation, as have the New York Board of Cooperative Educational Services (BOCES) units and the Modular Employable Skill-Based Training Program of the International Labor Organization, among others. Borrowing and adapting is efficient and desirable. Needless to say, wholesale borrowing is not likely to yield well integrated, sound programs.

In terms of what there is to borrow, while the information provided from task analyses can be of high calibre and useful for inferring specific knowledge and skills, it tells us little about worker characteristics which underlie performance on the tasks and is not very useful for drawing inferences about what causes or explains behavior. Task analyses alone are inadequate for operationally defining generic competencies. Valued behaviors such as being on time for work are either left out or are not analyzed in terms of necessary attitudes and motivation patterns that lead to this kind of behavior. Also, the supply of people with adequate theoretical understanding and conceptual skills to draw such inferences and develop training programs for these competencies are in extremely short supply (Tyler, 1980). As a result, the "affective domain," as it is called in the educational community, and complex cognitive skills are given very little attention in instructional programs.

Another problem in competency-based vocational education is the lack of articulation with regard to the general education components of the curriculum. In most cases, the general education component is not competency-based. By way of contrast, the general education portions of several liberal arts college programs carry most of the burden of developing generic competencies that cut across career areas. At present there are no incentives for such articulation at the high school level, especially in light of the fact that there is practically no information available on generic competency requirements of jobs.

One of the most promising aspects of competency-based vocational programs is their emphasis on skill clusters applicable to several jobs or occupational areas. Students who decide against auto mechanics, for example, can switch to plumbing and not have to repeat courses or modules already mastered which have bearing on both trades. Also, the cluster concept enables guidance personnel to individually tailor students's in-school and work-study programs. To improve capability in this area, a major study on transferable skills is underway at the National Center for Research on Vocational Education in Columbus, Ohio. Cornell University's Vocational Education Research Center has also received a grant to study basic skills in vocational education.

Despite limitations upon what competency-based vocational education is currently able to accomplish, improved relationships between employers and school systems and greater relevance of curricula to jobs are thought an adequate justification for the vast amount of energy currently being invested in program development. Congress had such objectives in mind in passing Public Law 94-482, Title I of the Education Amendments of 1976. The law requires that all vocational education programs receiving federal funds be evaluated every five years and that new ways be sought to evaluate student performance. Demand for guidance in developing performancebased measures is therefore increasing dramatically at the present time.

5. Credit for Experiential Learning

The history of external degree programs reaches back through the decades, and most external degree programs are traditional in terms of objectives, and modes of assessing student learning. In the early 1970s, however, some nontraditional external degree programs were developed in several states. Empire State College of the State University of New York and Thomas A. Edison College, the ninth state college of the New Jersey System, are outstanding examples of nontraditional external degree programs. A central focus of such programs is to create opportunities for adults to receive college credits for knowledge and skills acquired through experiences other than those directed by higher education institutions. Thus, the concerns of CBE and external degree programs are quite different. In the latter case, program developers must determine which life experiences are academically relevant and amenable to translation into college credits, while CBE programs seek to make academic experiences relevant to consequent life experiences.

Innovative external degree programs usually allow degree credits to be accumulated through a variety of channels, such as transfer of traditional college credits; credit through college equivalency examinations; individual assessment of collegelevel knowledge; evaluation of certificates or licenses; and evaluation of in-service training programs.

Most of these programs share information through the Cooperative Assessment of Experiential Learning Project (CAEL). The CAEL network has more than 200 members, including most of the competency-based programs in our sample. CAEL project offices located in Columbia, Maryland, represent a valuable source of written materials and contacts with member institutions which have developed performance-based and other nontraditional ways of assessing experiential learning and/or competence.

III. ASSESSING THE STATE OF PRACTICE

The CBE "movement," as may be gleaned from the summary account of trends during the last 10 years, is less a movement than an agglomeration of autonomous efforts to define outcomes which are consistent with program goals and the student populations for whom the goals are designed. There is much emphasis on linking curriculum to the demands of jobs and careers, and there are also many alternative routes for doing so. But, as pointed out in the section on vocational-technical education, it is not clear how institutions go about identifying the skills, abilities, and other personal qualities which underlie effective task performance, or, in other words, the generic levels of competency that are related to effective performance in adult roles.

With this as a consideration, it seemed logical for us to focus on competency-based undergraduate education in our research for this chapter, since the most ambitious attempts to identify generic competencies have taken place in undergraduate college programs. A good sample of these programs has been in existence long enough to have well-developed definitions of competence and operating assessment systems. Several of the programs had also reached the stage in their development where written documentation of their efforts was available to us. Moreover, extensive case studies were available on nine institutions through research supported by FIPSE (Grant, et al., 1979).

We attempted to sample the universe of competency-based undergraduate programs other than those for teachers. To locate institutions, we worked from a list identified by Gerald Grant, checking it against information available from FIPSE. We believe we were successful in identifying most of the fully implemented four-year undergraduate programs currently in existence. However, in light of state interest in competency-based vocational and technical education over the past few years, there are probably some fully implemented two-year competency-based postsecondary programs that our search failed to uncover. Data were also collected from the four competency-based professional school programs we were able to identify, and a vignette of the Antioch School of Law is included as an example of the model's application at that level.

In all, information on 19 competency-based programs (see Table 1) was collected in two manners. Key administrators of 13 programs were contacted by telephone and asked a series of prepared questions about selected aspects of their programs. In addition, six vignettes were prepared from case studies by Grant and his associates (Grant et al., 1979), from materials provided by the institutions, and from telephone inquiries to gain supplementary information. The vignettes describe five competency-based undergraduate programs and, for contrast, one competency-based law program. The draft vignettes were checked for accuracy and updated through additional correspondence and telephone conversations with officials at these schools. Additional sources of information for these reports are the personal familiarity of authors with particular programs and telephone contacts with officials at the National Institute of Education and FIPSE.

To provide a comparison with the CBE programs surveyed, it was necessary to obtain data on an additional sample of undergraduate programs to represent more traditional programs as a class. To this end, the authors surveyed 61 colleges and coworkers to obtain firsthand information on the way student assessment is conducted in other institutions. The survey of colleges through former students rather than school officials, was adopted because of the absence of institutional spokespersons at traditional schools who could or would answer the prepared questions, as well as time and resource constraints. Questions were therefore adapted for graduates of traditional programs. For our supplementary survey, 42 replies were received, including 8 reports of no undergraduate degree, 4 reports from graduates of nontraditional programs (for example, an external degree program), and 30 reports from graduates of traditional programs. In this last group of 30, only 2 schools (Harvard College and Boston University) were represented by multiple graduates. Their reports were combined for each school, yielding descriptions of 24 traditional undergraduate programs. This information must be taken as illustrative rather than definitive, because of the nonrepresentative selection of respondents, the unavoidable differences in the position of respondents in relation to the schools, and the form of the questions asked. Table 2 lists the institutions which comprise the traditional undergraduate program sample.

Brief Description of Selected CBE Programs

By way of introduction to the comparative analysis of how learning outcomes are defined and measured in traditional and CBE program, we thought it would be useful to present brief accounts of six CBE programs for which the most documentation was available to us. These vignettes are designed simply to convey basic information: the mission of the insitution, the kinds of students served, the major steps taken to define competence, and how the program assesses learning outcomes. The case studies recorded in Grant, et al. (1979) are additionally recommended as a way of developing an appreciation for the extraordinary effort that went into bringing these programs to their current stages of development.

1. ALVERNO COLLEGE (Milwaukee, Wisconsin)

Program Description. Alverno College is a small Catholic liberal arts college for women which has programs in nursing, management, music, and education. In three years of re-examining their institutional mission and goals, the faculty redefined

TABLE 1

CBE Programs Surveyed

*Alverno College
*Antioch School of Law
*College for Human Services
Delaware County Community College, Continuing Education
Florida International University, School of Education
*Florida State University School of Nursing
Governor's State College
*Kirkhof College of Grand Valley State Colleges
Mars Hill College
Metropolitan State, St. Paul
*Mt. Hood Community College, Registered Nurse Program
New Rochelle College, Teacher Education
Northwestern University, Master of Music Program
Our Lady of the Lake, San Antonio
Southern Illinois University School of Medicine
Sterling College
University of Massachusetts at Boston, College III
University of Minnesota, Pharmacy Program
University of Nebraska, Lincoln, University Studies Program

*Information gathered through case studies, program documentation, personal contacts, and telephone interviews. Data on other programs were gathered largely through personal contact and telephone interviews only.

TABLE 2

Traditional Programs Surveyed

Bentley College
Boston College
Boston State College
*Boston University
Chicago Theological Seminary
Dartmouth College
Florida State University
*Harvard College
Massachusetts Institute of Technology
Mount Holyoke College
Saint Paul's College
Smith College
SUNY Binghamton
SUNY Brockport
Syracuse University
Tompkins Cortland Community College
Tufts University
University of California, Irvine
University of Denver
University of Massachusetts, Amherst
University of Massachusetts, Boston
University of New Hampshire
University of Southern California
William and Mary College

*Four respondents each, responses combined liberal education in terms of the abilities it challenges students to develop. On the basis of their redefinition, they implemented an outcome-centered curriculum throughout the college in 1973 (Alverno College Faculty, 1976; Chickering, et al., 1977; Grant, et al., 1979; Cross, 1976).

Alverno identified eight interacting areas of ability, or "competencies": effective communications; analysis; problem solving; valuing in decision-making; effective social interaction; effectiveness in individual/environment relationships; responsible involvement in the contemporary world; and aesthetic responsiveness. Each competence area was set forth in a "learning sequence" of six levels, in which the student moves from developing discrete abilities in relatively isolated settings to integrating them in more dynamically open professional and community situations. Each student must demonstrate the first four basic levels in each competence, and advanced, specialized levels in competence areas appropriate to her major. Most students complete the basic levels during their first two years and the specialized levels during the last two years of their programs.

Competency Definition. As a starting point in competence definition at Alverno, each discipline presented its goals for students in a year-long series of faculty institutes. These goals were then synthesized by the college's faculty into broad outcomes or "competence areas." The Academic Task Force, with regular feedback from continuing faculty institutes, then proceeded to elaborate each competence area into six levels of increasing complexity. Each faculty member then analyzed his or her course(s) in terms of competence levels and stated when, where, and in what ways the course contributes to those learning outcomes and how student achievement of the competence will be assessed. Competence Divisions (for each competence area) made up of instructors from various disciplines then refined the specifications for each level and conducted interdisciplinary peer reviews of course designs, teaching/learning strategies, and assessment methods. This process of faculty and Competence Division analysis and refinement is ongoing as new courses and activities are added and old ones are refined; and it is an iterative process reflecting what is learned from student assessments.

Competency Measurement and Validation. The primary purposes of assessment at Alverno are: (1) to diagnose strengths and weaknesses of students and help them tailor their learning experiences accordingly; (2) to develop students' ability to evaluate their own performance; (3) to certify student progress; and (4) to use information from student assessments for program improvement (Alverno College Faculty, 1979). In designing their assessment system, staff of the college searched widely in business and industry, and in the education and research communities to learn what was available and suitable for adaptation. For example, some of the measures now in use were (e.g., the use of an audio-visual studio) adapted from assessment centers (Moses & Byham, 1977). Most of the assessments, however, were developed at the college by faculty and Alverno's assessment center staff, often with assistance from students. Assessment then occurs in a formal assessment center setting as well as in field settings or traditional classrooms. Measures themselves may be traditional (e.g., essay exams) or nontraditional.

Assessment techniques, like the competence definitions, are continuously evaluated and redesigned. The measures are generic, developmental, and holistic. Assessment criteria and techniques are explicit and shared among assessors. Results are presented to students immediately and in detail during structured feedback sessions. Special assessors are trained by the appropriate Competence Division of the College. Faculty are also trained in assessment design, and they and other assessors participate in brush-up sessions, updating, or special training workshops. Faculty, student peers or advanced students, alumnae, and members of the business or professional community are all part of the assessment system, and some of the assessments involve the use of jury panels composed of some combination of these assessors.

With special funding from NIE, Alverno has undertaken research to establish the validity of learning outcomes assessed at the college for later professional and personal success

(Alverno College OE, 1979). Alverno is validating the competencies and assessment techniques by adapting or developing validation techniques appropriate for use with their nontraditional assessment instruments (Friedman & Mentkowski, 1980). Student performance across and within competence areas is compared, and the relationship between student performance and external criterion measures is being examined.

As part of their validation studies, they are involved in comparing Alverno competencies with those demonstrated by outstanding professionals (Mentkowski, et al., 1980) and in following the future career paths of their graduates. They are also generating in-depth profiles of student perceptions of themselves and their development and analyzing the relationship of these to Alverno's learning process. Some special assessments of students' cognitive-development are also being made using measures designed for various levels of ego, moral, and intellectual development. These are then examined in light of external criterion measures for various analytic and interpersonal abilities.

2. COLLEGE FOR HUMAN SERVICES (New York City)

Program Description. The College for Human Services (CHS) was founded as the Women's Talent Corps in 1964 by Audrey Cohen (its current president) with funds from the U.S. Office of Economic Opportunity. By 1970, the College had changed from a one-year training program to a two-year associate of arts degree program preparing paraprofessionals, had admitted male students, and had changed its name to "College for Human Services." CHS faculty then concluded that its graduates were overqualified as paraprofessionals yet not fully prepared for professional roles. The College stopped admitting students between 1971 and 1974 and, with support from the Fund for the Improvement of Postsecondary Education, designed a three-year competency-based program to prepare professional human service workers. The term "human services" was coined at CHS to reflect a generic curriculum preparing graduates for roles across a broad range of social service functions and to reflect education and service delivery with an emphasis upon human needs as opposed to social control. Two new branches of the College have been opened, one in Broward County, Florida, and one in Alameda County, California. Both campuses award the bachelor's degree, and permission to grant the master's degree is pending in New York State.

One of the most dramatic aspects of CHS is the student body it serves. Its present class at the New York campus is composed as follows: all adults (youngest 22; oldest 55); 73% black; 19% Hispanic; 8% white. Slightly more than half are women, many of whom head households. When admitted to the College, about half of the students were on welfare; the other half were only slightly better off financially. Most of the students know firsthand the problems of the people they are trying to serve. Also, students' educational backgrounds vary considerably. Some have college degrees, others a smattering of college, some did not finish high school. Students are taught in the classroom and on the job, and are paid an average of $8,300 annually while they learn. First-year students are paid entirely from CETA funding awarded to the College. Second-year salaries come two-fifths from CHS and three-fifths from the student's field agency.

The CHS curriculum is arranged as a series of "constructive actions" to develop, practice, and demonstrate different facets of competence on the job with CHS faculty supervision. CHS's performance grid includes eight competencies, each reflecting five dimensions of performance: purpose; values; self and others; systems; and skills. The eight areas of competence referred to as "eight essential modes (crystals) of providing services" are:

1. Assume responsibility for lifelong learning
2. Develop professional relationships with citizens and coworkers
3. Work with others in groups
4. Function as a teacher
5. Function as a counselor
6. Function as a community liaison
7. Function as a supervisor
8. Function as a change agent

For a glimpse of the dynamic nature of the learning experiences of students at CHS and the flavor of the College, readers are referred to the article by John Grossman in American Education (1979).

Competency Definition. In 1971, CHS faculty organized into five committees on education, daycare, drug therapy, social work, and health professions, each of which included some persons from the agencies where students had been placed. Each committee attempted to define a standard of professional competence in that field, drawn from their own experience in supervision and observation, on the basis of job descriptions, policy statements by professional associations, licensing requirements, and performance.

In 1972, David McClelland and his associates at McBer and Company were retained by the College to conduct behavioral event interviews of 62 exemplary human service workers. The behavioral event interview is a specialized systematic data-gathering technique used as the basis for identifying and conceptualizing job competencies. McBer derived seven competencies:

1. Strong faith that human needs can and must be met
2. Ability to identify correctly the human problem
3. Ability to arrive at realistic, achievable goals in collaboration with clients
4. Imaginativeness in thinking of solutions to problems.
5. Persistence in pursuing solutions, often against hostile authorities
6. Ability to remain task-oriented under stress, hostility
7. Skills in getting interested parties to work together to arrive at common goals

This report became the basis of continued discussion among staff and was considered, along with other information when in 1974 the College developed a performance grid of eight competencies and five dimensions. From 1975 through the present, CHS staff have continued to refine the definitions using feedback from student field experiences.

Competency Measurement and Validation. The purposes of assessment at CHS are: (1) evaluation of student-practitioner performance, (2) improvement of service delivery, and (3) improvement of the education of professionals and their clients.

The major method developed by the college for assessing student performance is the "Constructive Action." Deborah Allen of CHS has defined "Constructive Action" as "the delivery of a human service in accordance with a plan. A Constructive Action has a beginning, a middle, and an end. The beginning is the working out of a plan answering to the goals and needs of a client (or several clients). The planners are the clients, the student-practitioner, the coordinator-teacher, and the agency supervisor. The middle is the carrying out of the plan, with such modifications as are appropriate. The end is the examination of the action, comparison of its actual results against the goals proposed at the outset. Satisfaction of the client-accompanied goals are, with an understanding and skill in self-sufficiency, the ultimate tests of the success of a Constructive Action, which is to be assessed at each stage by the four participants or groups of participants."

In the Constructive Action, strong emphasis is placed upon humane practice, educating clients for self-sufficiency, the education of professionals by clients, and shared control as a major outcome (Cohen, 1978).

Evidence of a Constructive Action is provided through written documents which must meet certain standards. In brief, these are:

- Full descriptions of agency, practitioners, clients, and other features of initial conditions that pertain to client's needs and resources for dealing with them
- Long- and short-term goals
- Strategies
- Detailed logs of all events making up the Constructive Action
- Justifications for any changes in goals
- Logs should add up to a demonstration of competence
- Evaluative analysis of logs with interpretations and conclusions
- Theory should be cited in proposal, logs and analysis
- Constructive Action must demonstrate knowledge and performance of a human service role
- Constructive Action should reflect active client involvement in planning and executing Constructive Action

Criteria for evaluating a Constructive Action came primarily from analyses of performance documents by former CHS students, drawing out the good points of superior Constructive Action and the deficiencies of inferior ones.

Other modes of assessment at the college include faculty evaluation of written work and evaluation of on-the-job performance by supervisors in the agencies and by clients (using a structured client questionnaire). Peers of student-practitioners also participate in some of the assessments.

Under a three-year grant from the National Institute of Mental Health that began in late 1979, the College will begin to move beyond content validity by conducting a major research project to test and refine its college/(community)agency collaborative model. Plans include a controlled study to compare performance of CHS student practitioners educated in accordance with their competency model with a more traditional student group in each of nine mental health agencies. The primary focus will be upon "citizen empowerment." Techniques to be used include structured interviews of the critical incident type and direct observations of behavior in both agency and college settings.

3. KIRKHOF COLLEGE,* GRAND VALLEY STATE COLLEGES (Allendale, Michigan)

Program Description. Kirkhof College, founded in 1973, was the nation's first totally self-paced college. Currently, self-pacing is just one of several instructional models and makes up 25% of all instruction (see Gamson 1979 for history of this idea of Kirkhof). The college specializes in career-oriented programs such as advertising, hospitality, and occupational safety. Kirkhof students may also pursue majors in other Grand Valley colleges. In the period before 1977, individual faculty were heavily engaged in developing discrete learning modules for students, most of whom majored in the humanities. Since the implementation of a competency-based model at Kirkhof in 1977, faculty attention was turned from learning objectives for course modules to defining generic competencies that cut across vocational and liberal arts areas. Only the general education portion of the curriculum, however, is competency-based. Ninety percent (90%) of Kirkhof's present students have a vocational major leading to a B.A. or B.S. degree.

---

* Kirkhof College was formerly called "College IV."

Thus far, 25 graduates of Kirkhof have fulfilled its competency-based general education requirements in addition to course and grade requirements in their majors. The required generic liberal arts competencies that faculty of Kirkhof believe to be necessary for successful personal lives and professional careers are:

1. Communication: (a) fundamentals of reading, writing, listening, and speaking; (b) composition of messages; (c) research and documentation; and (d) speech communication.
2. Problem Solving: (a) critical thinking; (b) effective problem solving; and (c) formal inquiry.
3. Social Interaction: (a) interpersonal interaction; and (b) organization and management.
4. Value Clarification: (a) identifying values; and (b) values and social conflict.
5. Application of Basic Concepts: (a) concepts in the natural world; (b) concepts in the social world; (c) concepts in the world of ideas; and (d) quantitative applications.

Competency Definition. Due to the college's interest in self-paced learning, during the period between 1973 and 1977 faculty energy was focused upon developing learning objectives for discrete modules and learning packages. In 1975, a Community Council was set up which included all full-time faculty, the dean, the assistant dean, and student representatives. The Council clarified and developed goals for the college and its students. These goal statements represented the beginnings of concentrated efforts to define a competency-based curriculum. Faculty were convened for a summer workshop in 1975 at which time they agreed upon generic competencies for general education. Faculty agreed to develop a competency-based program over the following two years. Products would be reviewed by the Curriculum Committee and voted on by the Community Council.

In the period 1975-1977, the communications competence area was developed as a prototype. Two faculty members (expert in oral and written communications) developed general statements of learning outcomes in the communications areas. A faculty member developed critical skills competence definitions. The process he used involved other faculty in specifying the generic skills required by their disciplines that related to various dimensions of critical thinking. Faculty consulted other competency-based education practitioners (including staff at Alverno) and adapted definitions or partial definitions from their programs where appropriate. From 1977 through the present, all faculty at the College have been involved in an ongoing process to develop definitions and measures of competencies.

Competency Measurement and Validation. There is a dual system of assessment at Kirkhof. Students may fulfill competency requirements through assessments within courses or through free-standing assessments. Jury panel assessments are conducted at selected points in the program and as an exit battery. These panels are composed of faculty members: Community professionals and practitioners sit on the juries for the exit battery. The College is now moving toward reducing the use of jury-panels to a midway checkpoint and to the exit battery.

The purposes of assessment at Kirkhof are: (1) demonstration of the skills associated with a liberal arts education; (2) evaluating student progress; (3) prescribing appropriate learning experiences; and (4) formative program evaluation.

The primary mode of assessment is by faculty constructed traditional paper-and-pencil tests. However, where appropriate, audio-tapes are used. Faculty are engaged in a continuing search for appropriate, cost-effective assessment techniques. Assessments occur in classrooms and in the testing center.

The exit battery conducted by a jury panel focuses upon "formal inquiry." This generic focus encompasses contributing competency areas of written communication, speech communication, critical thinking, values, research and inquiry, methodology and appropriate knowledge in a content area.

No validation research has been undertaken thus far and none is planned for the near future. Validation is, however, a goal of the college.

4. FLORIDA STATE UNIVERSITY CURRICULUM OF ATTAINMENTS (COA) (Tallahassee, Florida)

Program Description. Three pilot competency-based programs were implemented at Florida State University in 1974: biology (B.S.), nursing (B.S.), and urban and regional planning (M.A.). Additional competency components of B.A. programs were begun in 1975 in religion, music education, counseling, psychology, and theatre. The COA project defined mastery standards for degrees by talking with practitioners and potential employers of graduates. These standards were then used as a basis for defining new learning experiences outside the classroom, developing self-paced learning modules, and developing new assessment procedures. Faculty were used in nontraditional ways as mentors and tutors of students. By the end of the 1967-77 school year when programs began to atrophy, 232 students had been involved in the COA for some parts of their academic programs.

Except for nursing, competency-based programs or components of programs atrophied or vanished due primarily to staff turnover and lack of funds. The program for registered nurses serves nursing students whose basic education was obtained through a hospital diploma program or through a community college associate degree program. These registered nurse students enter the University program to earn a baccalaureate degree in nursing.

Competency Description. In 1974, faculty of the nursing program, at times in consultation with practitioners, produced learning modules presenting material on the role of the nurse, the profession of nursing, issues and dilemmas. In addition, abundant materials were produced on the basic sciences of physiology, anatomy, and social sciences relevant to the various settings for nurses. Modules emphasized textbook learning and performance in clinical or field settings. Each module included level competencies ultimately leading to terminal competencies. An example of a terminal competency is: "Use the Nursing Process systematically in providing nursing intervention to clients in any setting."

Within the University, any history or recollection of how competencies were defined in the other programs disappeared as the programs themselves atrophied. Documentation of the evolution and demise of the various programs can, however, be found in Grant, et al. (1979).

Competence and Measurement Validation. Faculty and nurse-practitioners assess nursing students in courses and at field sites using a variety of approaches. Jury panels are used for some of the assessments and students demonstrate achievement through simulation and actual client care situations. Many of the objectives of the program are assessed by traditional paper-and-pencil tests constructed by individual faculty members.

The program has no immediate plans for validation studies, although the institutional research department at FSU is prepared to assist in such research should funds become available.

5. MT. HOOD COMMUNITY COLLEGE, REGISTERED NURSE PROGRAM
(Gresham, Oregon)

Program Description. The competency-based Registered Nurse Program, culminating in the Associate of Arts degree, was founded at Mt. Hood Community College in 1970 by the late Diana Dean, Coordinator of Nursing Programs, and five experienced nursing faculty members. The Registered Nurse program replaced a Licensed Practical Nurse Program. The new program opened in 1971 with 30 students. Of 17 traditional nursing courses offered, three became competency-based in 1971 and an additional three in 1972. The program was fully competency-based by 1974.

The ideal for curriculum development was to precisely specify and assess behaviors in tasks customarily assigned to the nurse. The program at Mt. Hood has been innovative primarily in the sense that outcomes have been made highly explicit, ideas from behavioral science have been adapted, and students can proceed at their own pace. The program has undergone several revisions as staff has changed (Oleson, 1979).

Competency Identification. Through ongoing meetings and individual efforts, a team of five faculty members and the nursing program coordinator described nursing behaviors as competencies (e.g., taking temperatures; interacting supportively with patients). The team divided competencies into concepts and skills and focused on what they wanted students to write, to say, and to do. This effort culminated in the "write-say-do" model for curriculum, clinical labs, and field experience. The model is highly explicit and includes both generic and specific competencies.

As a second phrase, a twelve-person patient panel used Lindsley's rules for writing competence statements to develop 225 statements representing categories of nursing competence as viewed by patients.

The process of competency identification proceeds from the concept to be learned to the nursing behaviors which best identify the learning that has occurred. For example:

> "Concept: Hazards of Immobility (ROM, Body Mechanics and Positioning)"
>
> DO 28: Using appropriate methods, apply restraints as necessary to protect patient from injury.
>
> DO 29: Explain and validate the understanding of the patient and family as to the rationale for use and benefit from any restraining device.

Competency statements are also derived through faculty consultation and literature review. In addition, every effort is made to obtain feedback from students involved with the course competencies as to relevance and importance to their learning process as they perceive it.

Competency Measurement and Validation. Prior to 1975, competency assessment was thoroughly structured and formalized into what the college called the "Precise Behavioral Management

System." Its purposes were: (1) fine-grained monitoring of student progress: (2) evaluation of the impact of field experience programs on student performance: and (3) some faculty self-evaluation with respect to impact on student performance. Performance was assessed in practice labs, classrooms, and hospital settings where students were required to demonstrate skills. These skills were evaluated using highly explicit criteria shared among raters, including accuracy, speed and frequency of exhibition (referred to as the "Count System"). Course objectives were evaluated primarily through traditional written tests and recitations. Additional competency assessments occurred during the first and second quarter of the program and at the end of the second year. These assessments required students and faculty to record performance on "Program Product Behaviors" daily and quarterly on logarithmic paper. Data was stored in a computer program for retrieval and analysis.

A major revamping of assessment occurred in 1975 (see Oleson, 1979, for a discussion of the reasons for the changes). Grades were separated from the Count System. Classes teaching theory are evaluated in a traditional manner, taking into account student participation in class, and written work such mid-terms, final examinations and short, timed quizzes. The "Precise Behavioral Management System" and "Program Product Behaviors" terminology was dropped. Timing and accelerating frequency of skill exhibition were dropped as criteria in the Count System. Instead, performance in laboratory and clinical settings is now described in the following more generic terms: assessment, safety, efficiency, personalization, professionalism, teaching, intervention, and comfort (ASCEPTIC). Student performance is evaluated by narrative, critical incident reporting, and final scores are determined by percentages from 100 through 60 and below, translated into A through F.

The college has no plans for further validation research at the present time.

6. ANTIOCH SCHOOL OF LAW (Washington, D.C.)

Program Description. The Antioch School of Law (ASL), founded in 1972, emerged from an affiliation between the Urban Law Institute and Antioch University. ASL's goal is to broaden the range of legal services available to poor people in the nation through both the delivery of direct legal services by attorney-professors through the law firm and the expansion of the range of individuals with legal training by means of a liberal admissions policy for the three-year juris doctor degree.

The juris doctor program is an intensive three-year program that is academic and clinical in nature. It is in the clinical component of the curriculum that the competency-based program has evolved. There are two components of ASL: the law firm, which handles approximately 2,000 cases per year, and the law school. ASL's purpose is to train students to be effective client advocates through traditional academic courses and clinical work in external, client service clinics, and in the law firm which serves poor people from the District of Columbia. The recently reorganized law firm is divided into two divisions: Public Law and Private Law, with criminal law components in each division. For graduation, students must have accumulated 62 academic credits and 30 clinical credits, which includes experience in client service clinics and a semester internship in local or federal government agencies. Clinical credits are achieved through satisfactory completion of six consecutive clinical rotations. Students are evaluated in clinics on six generic competencies:

1. oral competency
2. written competency
3. legal analysis competency
4. problem-solving competency
5. professional responsibility competency
6. practice management competency

Each general area of competence contains specific dimensions with criteria for evaluating each. The criteria tend to be possible errors of omission and commission.

Although not included with the case studies in the published book (Grant, et al. 1979), a case study by Nancy Jones (1978), provided in a report prepared for the Fund for the Improvement of Postsecondary Education, presents an excellent account of developments at ASL through 1977.

Competency Definition. From 1972 to 1974, faculty and consulting legal scholars worked on several problems such as how to structure the law firm, the content and sequencing of academic courses, and their mesh with the law firm. In 1974, with funding from three grants, the Competency-Based Task Force was formed at the law school. Its membership included the project directors of the grants, the co-deans of the law school and three consultants who had backgrounds in psychology and education. The consultants worked with faculty to develop an initial list of "lawyering skills." A faculty retreat was held, which yielded a consensus on specific skill areas and the use of specific language for describing the "lawyering" process.

In 1975, the Task Force employed several methods to gather data for use in refining and revising initial lists of competencies. To help guide development of the academic curriculum, seven traditional D.C. law schools were surveyed. Faculty members in these institutions were asked what criteria they used to evaluate their students. For design of the clinical component, a twelve-page questionnaire was administered to ASL faculty to determine, for example, distinctions between legal interviewing and interviewing conducted in other programs. Clinical faculty also developed scenarios that contained detailed descriptions of their lawyering activities as they served clients in the (then) three divisions of the law firm. The scenarios included descriptions of legal situations and contexts for legal issues and problems. Also, the college had developed "professional board examinations" which involved jury panel assessment of student performance by practicing attorneys. The professional boards were also used in refining the competencies. (The professional boards have since been abandoned due pimarily to cost saving considerations.)

In the period between 1975 and 1976, the identification and refinement process continued. The main product of this effort was the development of the Basic Clinic Progress Report, a form evaluating student performance on tasks in the law firm.

Prior to 1976, ASL's approach to curriculum and assessment design was to do task analysies which yielded primarily specific skills. Dissatisfied with this, the Task Force took a fresh look at the information that had been gathered and switched to identifying generic competencies of lawyers applicable across a broad range of tasks. As a result, they identified six major areas of competence, then defined specific competencies for each with criteria for evaluating them. This effort resulted in the publication: Catalog of Definitions of Generic Lawyering Competencies (Competency-Based Task Force, 1978).

Competency Measurement and Validation. The major purposes of assessment at ASL are: (1) to diagnose strengths and weaknesses of students and tailor their clinical experience accordingly; (2) to develop students' ability to evaluate their own performance; and (3) to use information from student assessments to refine competency objectives and other aspects of the program.

While, applications of the competency-based model for traditional courses are being explored, academic courses at ASL are evaluated almost exclusively in traditional ways. Our discussion of competency-based assessment pertains only to the clinical component of the curriculum.

For each of the six clinics that students must complete, supervisors must select a minimum of two tasks to closely evaluate in accordance with the competency-model's criteria, using the Basic Clinic Report form. The tasks selected must be recognized to be of central importance for a field of law handled in the clinic and must be ones for which the supervisor can observe products or performance. Different clinics identify different tasks. A six-point grading system is applied. Students must get a minimum rating of 4 out of 6 for each of the two tasks in order to pass the clinic. Records of student performance are then filed with the Registrar. In addition, some clinic faculty make use of audio-visual tapes and other nontraditional approaches to student assessment. The Catalog of Definitions of Generic Lawyering Competencies has also been used to identify criteria for evaluating the legal memoranda of beginning students. For these evaluations, the criteria have also been set out in accordance with a point system with letter grade equivalencies.

Student assessment occurs in courtroom settings, classrooms, law firms, external clinics, and in internship settings. Performance is evaluated by student peers and internship supervisors as well as by the academic and clinical faculty.

ASL has established content validity for its assessment of lawyers entering practice where they will act as lawyer/

advocates for poor people but has not undertaken construct, concurrent or predictive validation studies. Nonetheless, a major effort is being made to assure an adequate data base will exist should resources become available to undertake such studies.

Data are currently being converted for entry into the school's computer system and will be used for evaluating relationships between professional outcomes (e.g., bar status and employment) and competencies, learning, experiences, admissions variables, academic performance, and clinical performance. Data on the past three graduating classes will be computerized by Spring of 1980.

To date ASL faculty have informally attempted to follow-up graduates and use this information for program adjustments. Also, a major reorganization of ASL occurred as a result of an extensive self-study where staff examined goals and how well extant programs related to them.

IV. THE SOURCES AND NATURE OF COMPETENCE GOALS

The first group of questions to be addressed in our survey of CBE institutions concern the source and nature of the competencies or education goals of the institution. Where did these goals come from, and how were they defined and described?

Educational goals have a variety of origins. Sometimes they are looked upon as self-evident goals, such as an appreciation of beauty or maturity of character. They may be conceived as "relevant" to later life, but often the relevance rests largely on tradition and is neither fully specified nor convincingly demonstrated. For example, ever since Plato's Republic, educators have believed that qualities such as critical thinking and conceptual skill were important for effective citizenship. The terms "critical thinking," and "effective citizenship," however, are usually not defined in any rigorous way, and the assumed relationship between them rests on faith, tradition, or plausibility rather than upon any careful demonstration that people who think more critically are actually more effective citizens. For convenience, goals of this kind can be labeled *a priori*. (Goals specified by the law or by sponsoring groups and "authorities" such as charities or boards of trustees, can also be included here.) These can be contrasted to goals for which the link to later life is either obvious or else established by careful empirical demonstration, a category of goals which we shall call "demonstrated." An example of a case where the demonstrated relevance of a competency is obvious would be assuming knowledge of navigation and engine mechanics to be relevant to flying an airplane. Cases where the demonstration is empirically established, but is not necessarily "obvious," are a little harder to identify. The ideal case would be training to goals or competencies which have been established as relevant by means of an analysis of the skills and competencies which actually distinguish superior performers in a given role or occupation from average, and even ineffective performers. Examples close to this ideal include educational programs designed to increase achievement motivation on the basis of empirical evidence linking achievement motivation to success in entrepreneurial roles (McClelland & Winter, 1969) and the Leadership and Management Education and Training programs recently developed for use by the United States Navy (Winter, 1979).

A second question about educational goals or competencies is whether they are generic or specific in nature. Specific goals or competencies are applicable to one or a few situations in later life; generic goals are broader, more abstract, and applicable to many situations and to a variety of adult roles. In principle, the source (a priori vs. demonstrated) and the nature (generic vs. specific) of competencies or goals are independent domains. We can argue that the dichotomy reflects, at least to some degree, a value basis for making judgments about the goals of training programs. To the extent one believes in empirical evidence for goal relevance rather than received tradition, one should prefer to have goals that are "demonstrated." If an institution has largely goals of this type (i.e., connected in either an obvious or a demonstrated way to the desired performances after college) it is, arguably, at least more efficient because it is one step closer to providing training relevant to later life. No such value judgments, however, can be attached to the distinction between generic and specific competencies. Generic goals might be seen as more "efficient" because they are broader and apply to many types of situation. On the other hand, if a goal is important even for only one job or situation, there is nothing wrong with its being specific. Future neurosurgeons and airline pilots must be taught a great number of highly specific skills and competencies, and many lives depend upon their success in learning them. Clearly, specific skill and competency goals should not be avoided because they are not generic enough to be useful for multiple purposes.

What was the source of educational competence goals at each of the institutions studied in this report, and how specific is the language in which these goals are phrased? By way of answering these questions, we will refer to a smaller subset of institutions as reference points for our discussion of the results of our survey. From liberal arts programs, we include Alverno College, where the CBE program includes the entire institution, and Kirkhof College of Grand Valley State Colleges, where the CBE program exists as a relatively small part of a larger, more traditional institution. From professional schools whose curricula were developed from a competency-based perspective we consider the Antioch School of Law and the Southern Illinois University School of Medicine. Finally, from schools preparing students for nontraditional or innovative professions we include the College for Human Services. To begin our discussion, we want to consider a member of the group of traditional liberal arts institutions as a touchstone for our discussion of the extent to which CBE programs differ from programs developed through more orthodox approaches. Because of its standing among this latter group and the recent notoriety surrounding the recasting of its general education program, we have chosen Harvard College to represent the traditional programs.

Harvard College

At Harvard, under the leadership of Dean Henry Rosovsky, the faculty committee that created the core curriculum (Report on the Core Curriculum, 1978) conceptualized the ultimate goal of undergraduate education (the later performance for which students are educated) as the simple state of "being an educated person." They proceeded to ask themselves:

> "What does our faculty and our university mean when we welcome a student at graduation 'to the company of educated men and women'? What are the characteristics of an educated person and how can curricular requirements help to foster these?"

In response, the committee elaborated five goals for undergraduate education, which they felt characterized educated persons, or ought to characterize them, in "the late twentieth century." Those goals were:

(1) Thinking and writing clearly;

(2) A critical appreciation of ways in which we gain knowledge;

(3) Knowledge of other cultures and times;

(4) Understanding and experience regarding moral problems; and

(5) Depth in some field of knowledge.

These five characteristics are defended on purely _a priori_ and normative grounds. There was no attempt to marshall evidence that any or all of these five characteristics in fact do characterize "educated people . . . in the late twentieth century." For example, the committee argued that, "A crucial difference between the educated and the uneducated is the extent to which one's life experience is viewed in wider contexts." They cited no evidence that education as such (compared to media exposure, factory experience, or some other confounding variable such as travel) may, in fact, lead people to view their life experience in wider contexts. In like manner, the committee confidently suggested that: "It may well be that the most significant quality in educated persons is the informed judgment which enables them to make discriminating moral choices." Such a conclusion would be contested by a wide variety of people, ranging from critics of the role of America's "brightest and best" educated leaders in the Vietnam war, back to Clarence Darrow's famous defense of Richard Loeb and Nathan Leopold as crippled by "the disparity between . . . intellectual development and . . . emotional life, . . . an emotional poverty, compared with an intellectual wealth" (Higdon, 1975, p. 225). While the argument by the Harvard committee is authoritatively persuasive, it is nevertheless an "armchair" argument. The five elements of "being an educated person" are defined at a very high level of abstraction, and no further breakdown is given in the course of the report. One can, therefore, describe the Harvard Core Curriculum committee proposals unambiguously as _a priori_ and generic.

Alverno College

Alverno College is an outstanding example of a liberal arts college thoroughly committed to the competency perspective. After a good deal of faculty discussion and experimentation, eight broad, generic competence areas were formulated, and within each broad area, six levels of progression were delineated. For example, in the "communication skills" area, the first level is to "identify own strengths and weaknesses as initiator and responder" in six different types of communications, while the sixth and highest level is to "communicate effectively through coordinated use of three different [technologically advanced communications] media." Each level of each competence area further specifies indicators of competent performance at that level. For example, the specifications for level 3 fill six single-spaced pages. The criteria for writing at level three include eight standards, one of which (clear, effective form) is in turn broken down into nine components.

Compared to the statements of goals and outcomes made by almost all traditonal liberal arts colleges, Alverno's generic competency goals are distinctive for their extraordinarily clear and specific language. Their source, on the other hand, seems to be traditional; i.e., faculty discussion, reflection upon received theories about liberal education, and the logic of pedagogical experience.

Alverno College is committed to educating women for adaptive management competence under conditions of rapid social, community, and personal change. While all of its competency goals are plausibly related to the outcome goal, the linkage rests on rational argument and _a priori_ analysis rather than upon an examination of the actual traits and skills of women who are demonstrably successful in these ways.

Kirkhof College

After several years of experimentation (see Grant, et al., 1979, Chapter 12), Kirkhof College (College IV) at Grand Valley State Colleges evolved a competency-based liberal arts program that is in some respects a variant of Alverno's. It specifies five areas of competence, but apparently only a single level within each area. Although the introductory guide to the program claims that these five areas encompass "the abilities or skills which allow [college graduates] to function effectively in both personal and professional life," and that "this degree design is based on actual community needs," no evidence is presented to support the claims. In fact, the competencies emerged from faculty discussion, previous experiences developing programs in "communications" and "critical thinking" competencies, and input from other programs such as those at Alverno and Bowling Green State University. Though the language describing these competencies is more general than that of Alverno and more specific than that of traditional liberal arts colleges, the competency goals of Kirkhof College are both generic and *a priori*.

One preliminary conclusion is that liberal arts colleges, whether traditional or competency-based, derive their competence goals largely from the authority of tradition, faculty speculation and discussion, and ongoing pedagogical experience; all sources that we have grouped together as "*a priori* analysis." Most liberal arts colleges claim that they are preparing students to perform competently in their future personal and professional lives, but few colleges have actually gathered evidence about the actual skills, traits, or other competencies associated with those desired long-range goals as a source of data for curriculum design.

Southern Illinois University School of Medicine

One might think that the goals of professional schools would be derived differently, with the criteria of ultimate "success" presumably more clearly defined and more carefully researched. Our analysis, however, shows that this is only partly true. Take as an example the Southern Illinois University School of Medicine which is now completely competency-based. According to principles adopted in 1972, "The focus of development and implementation of the curriculum will be on what students should learn as opposed to what teachers shall present" (Silber, et al., 1978, p. 474). Presumably what students should learn is the knowledge, skills, attitudes, and values that characterize successful doctors (as compared to the general public or, by a much more sophisticated derivation, as compared with less successful doctors). The actual formulation and continued refinement of objectives have been based on various data sources, including other medical schools, consultants, prior faculty experience, and specialty board recommendations (Silber, et al., 1978,). These sources are reminiscent of those used in other _a priori_ approaches. Nonetheless, the school's competency objectives are broken down and formulated in great detail. There are more than 8,000 stated objectives in the curriculum, most of which describe specific knowledge and skills required in medical practice. Consequently, we would describe the SIU Medical School program as having highly specific, although _a priori_, competency goals.

Antioch School of Law

In contrast with the other four institutions cited thus far, the Antioch School of Law performed an analytic study of the lawyer's job to supplement the traditional goals of the law curriculum. Antioch's law firm faculty were asked to fully describe their activities in "clinical" (i.e., practice) settings. In addition to the specific knowledge and skills derived through rational, *a priori* analysis (similar to that of the SIU School of Medicine), six generic lawyering competencies were formulated, broken down, and defined in a 63-page catalogue (Competency-Based Task Force, 1978). As Cort and Sammons wrote, "The Antioch Competency-Based Task Force model of lawyering competencies derives from the concept of the lawyer as communicator, legal analyst, problem solver, practice manager, and responsible professional" (1978, p. 5). The competency goals of the Antioch clinical component, therefore, are a mixture of generic and specific statements based on job analysis of what lawyers actually do in this type of firm; whereas the academic portion of the curriculum is based entirely on *a priori* analysis.

College for Human Services

The College for Human Services was founded in an attempt to define and train people for a new profession. Professional training was to be focused on "humane performance rather than simply academic knowledge." After a year of discussion about the "generic competencies" of the profession of human service worker, the College retained Dailey and McClelland (1974) to develop, along lines articulated by McClelland (1973), an empirically-based model of competent performance in human service jobs. Thematic analysis of critical-incident interviews with 62 incumbents who were nominated as outstanding human services workers yielded seven competency elements:

(1) Strong faith that human needs can and must be met;

(2) Ability to identify correctly the human problem;

(3) Ability to set realistic, achievable goals in collaboration with clients;

(4) Imaginativeness in thinking of solutions to problems;

(5) Persistence;

(6) Ability to remain task-oriented under stress; and

(7) Skill at gathering interested parties to work together to arrive at common goals.

It is difficult to know how much the empirical study of Dailey and McClelland actually affected the final form of the curriculum. One or two of the eight competencies seem to be related to the seven characteristics of actual, superior human service workers as identified in that study. The comparable goals also tend to be the sort of goals one would have expected to emerge from a purely a priori analysis. Ultimately, the college combined information from empirical research and a priori analysis to develop a performance grid containing eight generic competencies, each with five dimensions or facets for "constructive action" toward competence.

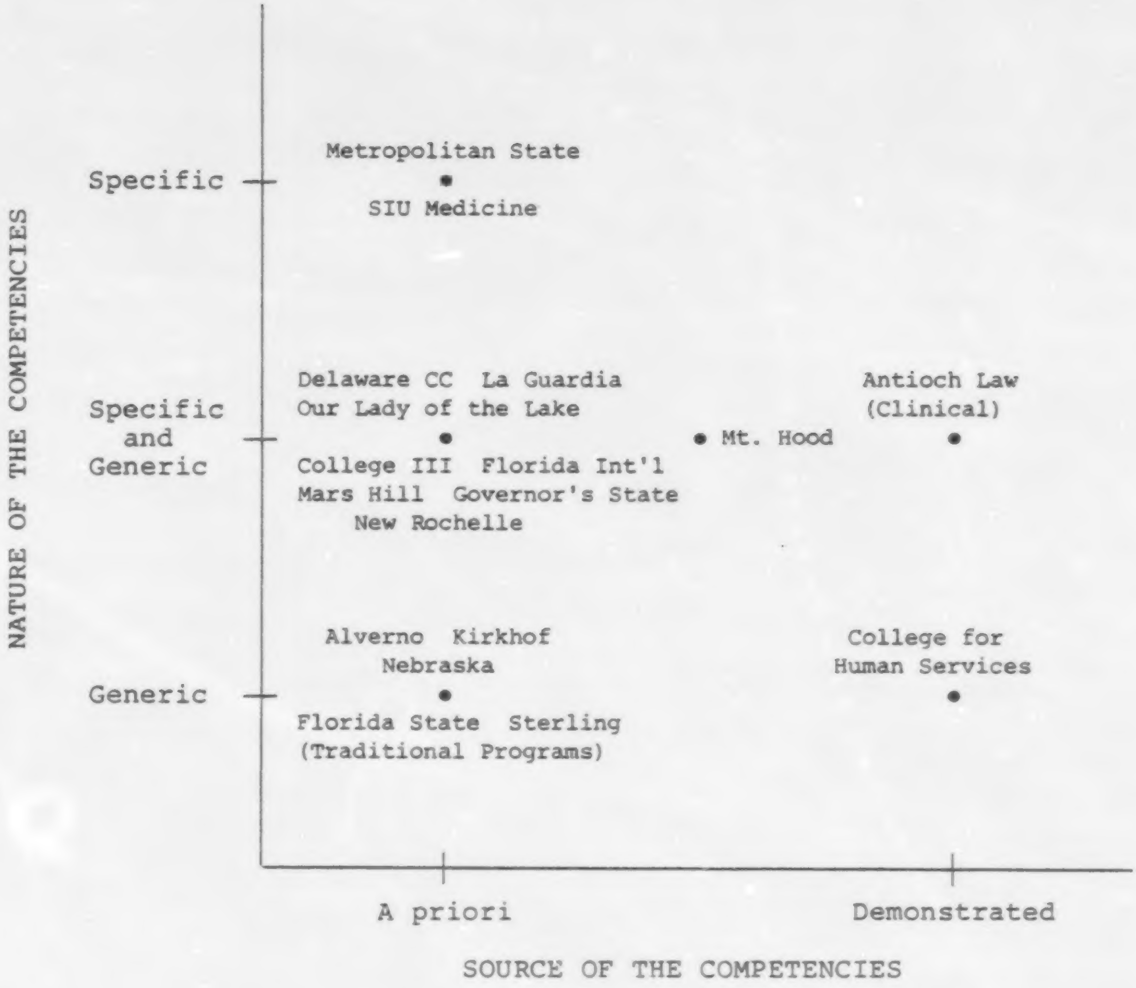
Figure 1 presents a plot of the CBE programs in our survey according to the source and nature of the institutions' competency goals as previously discussed. Traditional programs are represented here also, but there being little or no variation among these institutions (i.e., goals are almost invariably *a priori* and generic), they are grouped together in the figure.

As can readily be seen, the major source of variation among competency goals derives from whether the competencies emphasized are generic or specific. But for all the differences in philosophy, tradition, and rhetoric, most of the CBE institutions surveyed are close to the traditonal liberal arts model in terms of the source of the competencies. Though the goals are often different in content and elaborated in more specific terms, these goals were largely derived on an *a priori* basis and have not been demonstrated to be related to later life outcomes.

Educational programs focusing on particular professions tend to mix specific competency goals in with the generic to a greater extent than do strict liberal arts programs: But even these mixed programs have not taken into account dimensions of competence which have been shown to distinguish effective and ineffective performance in professions. This is not to say that the competencies identified by CBE programs, or any other programs for that matter, are invalid; rather that the demonstration of their relationship to later performance was not a

FIGURE 1

Types of Educational Competency Goals Classified by Source and Nature



consideration in the original curriculum formulations. A separate issue, then, is whether, despite this, CBE institutions are now working toward demonstrations of the relevance of competency goals to late life outcomes.

As our major focus is the measurement of competency attainment in higher education, our attention now turns to a definition of different aspects of competency-attainment measures and to the degree to which competency measurement has been used to demonstrate the relevance of institutional goals.

The Practice of Assessing Competency Attainment

Once a set of educational goals has been selected, articulated, and supported by a curriculum and pedagogical approaches, student performance is to be assessed and evaluated with respect to the goals. Assessment and evaluation have generated perhaps a wider array of philosophies and practices, and a greater amount of emotion and controversy, than most other issues in higher education.

Whether based on a multiple-choice tests, a research thesis, or a personal journal; whether graded with an A-F scale, a credit/no credit dichotomy, or a lengthy "jury" opinion, assessment and evaluation always occur. Some procedures for measuring competency attainment are used, and some standard of evaluation are employed. Professors who did not evaluate in some way would not be thought of as living up to their responsibilities. Equally, faculty members who gave every student exactly the same grade, without determining whether the students' performances were in fact equivalent, would also be thought remiss in their duties. Whatever the implications of assessment results and evaluation of competency attainment for later credentialing and tracking, they do answer the elementary epistomological question that is basic to teaching: "How do we know that a student has learned that which was taught?"

The issue of assessment is particularly important because of the emphasis of CBE programs on the demonstration of a student's ability to apply acquired knowledge: Such achievement is assessed by nontraditional means. To supplement multiple-choice or essay examinations and term papers, students are asked to demonstrate behaviors which are thought to be important to the later milieu or roles for which they are being educated. This section of the chapter, therefore, deals with the means of competency assessment and the degree to which they differ from those of traditional programs.

We began our survey of CBE programs with several aspects of the assessment process in mind. Our goal was to find out what we could about the practice of competency assessment without suggesting desired prior categories of response. Our data gathering procedures were informal and semi-structured telephone interviews, and we proceeded much as would an investigative reporter. Nevertheless, we focused in particular on four key issues about the nature of assessments used in CBE institutions:

1. The Nature of the Competencies Measured. Are the measures of specific competencies based on content knowledge or skills having limited use, or are they of generic competencies applicable to a broad array of tasks and across different fields? It is, of course, important to assess both types of competency in order to determine if students have learned what was expected of them. But the nature of the measure itself may confuse the distinction between specific and generic levels of competence. For example, consider an essay question, typical of a final examination in the psychology of personality designed to test for "critical thinking": "Compare and contrast the motivation theories of Freud and Maslow." Certainly this question taps a student's ability to remember and use factual knowledge about Freud and Maslow, a competency highly specific to the field of psychology. The question is tied to a particular subject matter, but it is easy to accept that the ability to marshal, organize, and use facts reflects a generic competency. In practice it is difficult to parcel out one aspect of competence from another, and attaching a single letter grade to the student's response further obscures the distinction.

More important from the viewpoint of those who seek to ascertain the meaning of the college degree is the issue of comparability of different means of assessing either the same or different competencies. Tying measures closely to content within disciplines makes it difficult to establish convergence at the generic level. For example, is a test of "critical thinking" for experimental social psychology measuring the same things as the physiological psychology critical thinking test, or the other tests custom-designed for use with macro-economics, German romantic poetry, or Wittgenstein's philosophy? And how do these measures at one institution compare with other, different ones at another institution? The greater the proliferation of outcome measures, the greater the work necessary to demonstrate convergence and equivalence (in order to establish fairness). It would appear to be necessary to standardize the sets of criteria used to judge possession of specific and generic competencies.

2. The Explicitness of the Assessment Criteria. For the sake of fairness as well as accuracy, the degree to which assessment criteria are spelled out and defined will determine the extent to which different evaluators make the same assessments of the same student performance. In psychometric language, this is called inter-rater reliability or agreement. Explicitness provides some assurance that, whatever the worth of the competency being assessed, the asessment process itself does not add error or bias. But there are two additional aspects of explicitness which are more important from an educational perspective. One is the explicitness of

performance expectations prior to assessment. The degree to which faculty expectations for what students are to accomplish in a course or a program might have a bearing on the worth of a final assessment in determining what a student has learned. If expectations are either unclear at the outset or at variance with the reality of the testing situation, the competency certification process can, at the very least, be deemed unfair. Teachers who set up different assessment expectations from those reflected in their final examinations can be criticized strongly on ethical grounds. Indeed, such practices might be susceptible to challenge in the courts, as has happened in the minimal competency testing area where tests were established without proper regard to that which was taught (Debra P. v. Turlington, 78 Civ. 892).

Another aspect of explicitness is whether or not specific assessment criteria are made public by assessors. A combination of detailed assessment criteria which assessors can share and clear expectations of student performance in terms of those criteria enables students to know the goals toward which they are studying and working, so that learning becomes more of a collaboration between student and professor. This combination of criterion attributes also permits students to question evaluations and faculty to defend them, thereby eliminating the specter of unfairness.

3. The Measurement of Competency Development. Measures of competency should have "educational validity," performance

should be shown to change as the result of educational programs and interventions. Students who complete the program should score higher than they did at the beginning of the program. This issue seems obvious, yet it is often neglected or treated casually in practice. Careful consideration of this issue first implies careful coordination between curriculum design and measurement of competence. We should try to test that which we have taught; and if we evaluate students' performance with respect to a certain criterion, then we should design our teaching to be sure that we emphasize that criterion, either explicitly or implicitly. Second, it implies the use of measures that are capable of showing change. Measures that appear to reflect "traits," such as intelligence or abilities thought of as stable, personality (if that is thought of as stable), or abstract stylistic traits, will not, therefore, be very useful for measuring the competence goals of higher education. In fact, because the canons of personality and ability testing require that these measures be stable or reliable (that subjects give the same answers on a second testing that they gave on the first testing), they have a built-in bias against showing educational change. When measures of this kind are used to evaluate students' performance or program effects, they tend to work against showing any change, not necessarily because there has been no change, but because the component "items" of these measures have been screened so that only the most "stable" (resistant to showing change) survive.

This discussion suggests that most traditional measures of ability, such as those used to screen students at entry, are not very useful as measures of learning or competence. Knowledge tests, per se, would obviously be expected to show change with education. What we do not know is whether nontraditional measures of ability, perhaps measures of generic competence used in CBE or other programs, are capable of showing change with education, thereby demonstrating, at least, in part, the educational worth of certain competencies.

4. The Validity of the Competency Measure. There are several perspectives from which one can judge the validity of a measure. Generally speaking, a valid measure is one which measures what it is supposed to. Test designers and users may approach validation of a measure with more or less rigor, ranging from an intuitive acceptance of validity to a rigorous statistical analysis of the measure's validity. If a competence measure appears (at face value) to measure what it is intended to measure, then it may be considered face valid. If upon careful logical review a test seems to sample and represent the domain it is intended to measure then it may be considered content valid. For example, if a competence measure is intended to assess whether a person has knowledge required for a job, it should measure an appropriate sample of topics in that knowledge base; if a measure is designed to determine whether an individual will respond competently to circumstances in a job, then the measure should include a representative subset of

situations for which a certain response is required for effective performance. The extent of assessed responses and the pattern of responses on a test of established content validity may be used to assess the competence weaknesses, and contradictions in a person's background or ability to perform a job competently. Construct validity is another logical (or rational) judgment on a test's ability to identify and measure a particular attribute of the test taker. Since it would be difficult, for example, to identify, sample, and defend a choice of test questions to reflect the domain of intelligence, the validity of an intelligence test is generally established by comparing the results of a new test with results of existing tests. If the results of a new test agree with accepted tests of the construct intelligence and disagree with tests of other constructs in predictable ways (e.g., unrelated to manual dexterity) then the test may be judged to be construct valid. Lastly, the most demanding and informative form of validation is known as criterion-related validity and is established by statistically correlating performance on the measure with later performance (criterion) which is presumed to be influenced by or related to the competence measured.

The aspect of validity which interests us most in CBE is criterion-related validity, in this case whether a measure of educational competence has demonstrated relevance to performance in later life. Here the issue is not whether the measure is relevant, but the basis for the claim and the kind of evidence brought forward to support it. All educators probably believe that their measures of evaluation are "relevant" to later life: The question is whether this belief rests on faith or on some kind of demonstrated evidence about the measures. There are typically two ways to determine evidence for this belief: Either examine whether people who are more effective in their careers or in life perform better on the measures than their less effective counterparts (concurrent validity), or see the extent to which students who score higher on a given measure or set of measures go on to be more effective in achieving work or other life outcomes than students who scored lower on the same measures (predictive validity). At the outset, note that predictive validation is a highly unlikely process for a college to undertake, owing to the time and cost involved before results can be substantiated.

Having specified the four issues we attempted to address in our survey, before proceeding to our findings, it may serve us well to examine the typical, traditional means of testing for student competence ordinarily employed in undergraduate college programs. We return to our example of the essay examination to see how many teachers go about evaluating students and discuss this mode of assessment in light of the four issues of interest. Suppose the essay question, "compare and contrast the motivation theories of Freud and Maslow," were the subject of a final examination in psychology.

Scenario: Late at night, the day before the final grades are due at the registrar's office, the professor, plowing through the exams, picks up the next bluebook and reads through it. The professor's thoughts run something like this:

> Well, this student says most of the right things about Freud and Maslow. The facts are pretty good. the organization is okay, but it rambles a lot, and that whole first page of discourse on Freud being "linear" is off the point. Well, all the facts are there, but the organization just isn't up to "A" standards; it's not creative and crisp. But it's better than the usual "B" rehash of what I lectured about in class. I guess I'll give it a B+." I wonder, though, is it really worth an "A-?" Isn't this as good as that exam I gave an "A-" to an hour ago? Oh my, it's so late and I have so many more exams to do. "B+" it is then.

How can this familiar measure of educational competence attainment be characterized in light of our four issues? First, what is the competence being measured? Is it factual knowledge of Freud and Maslow? Apparently so, but there also seem to be some other implicit competencies involving organization, synthesis, and writing style, at least so far as the threshold between an "A" and a "B" goes. Is the exam generic? It is tied to particular subject matter (Freud and Maslow), but it is easy to accept that the ability to marshal, organize, and use facts reflects a generic competency. A student's answer to this question would be of about the same quality as his or her answer to any other similar essay question, assuming some minimal mastery of the facts in question.

Are the assessment criteria explicit? Yes and no. The professor can usually defend his or her grading to a disappointed student, but without being very explicit about what is expected, especially in the area of organization of facts and clarity of writing. Would colleagues agree with the professor's evaluation? That is an interesting question, but normally faculty members don't grade each other's exams. One cannot determine how clearly expectations were specified in advance of the examination. Presumably the question might not have caused great surprise given the scope of the course; it might even have been given out in advance. But most likely, the expectations of this faculty member about what it would take to get an "A-" on the test were largely unarticulated.

Does this measure show change as a result of education? The professor presumes it does; but in fact we do not usually know. Few professors have ever tried giving the final exam on the first day of class or devising some method for assessing educational competence before and after a course. Presumably a student who hadn't learned anything about Freud and Maslow would not write a very good answer, therefore the "facts" aspect of the measure does reflect the experience of the course. On the other hand, it is by no means clear that the "use and organization of facts" or "clarity of writing" aspect of the measure (the competency relevant to the "A-" / "B+"

borderline area) has anything to do with the personality course, or that it would show change as a result of the course (controlling somehow for factual knowledge). Within the context of a single course, other than a composition course, most professors probably treat writing style as a stable trait rather than as something that can be changed, even if they don't grade that way.

What about the validity of the measure? Most people would agree that the question is face valid: it "looks like" it measures students' ability to remember and then use factual knowledge about Freud and Maslow. Certainly in that respect it is a better measure than question such as: "The unconscious played a more important role in Freud's theory of motivation than it did in Maslow's. True or False?" But what about relevance to later life performance? Most professors believe that their assessments of student performance through exam grades (and the letters of recommendation that are based on these and other grades) are full of importance and relevance for students' later lives. How can someone who can't answer the Freud and Maslow question turn out a success later in politics, business, or law, let alone as a professor of psychology! Yet there is not very much evidence that college grades have any strong relationship to competence in later life (see Heath, 1977; McClelland, 1973). More important, professors do not usually have the ability to, resources for, or interest in gathering such evidence systematically.

In summary, the traditional method of measuring educational competencies, typified by the preceding hypothetical but realistic example of exam grading, is to use a series of private judgments backed by undemonstrated personal confidence about later life relevance. The strongest characteristic of this method is its face validity. Almost by definition, faculty believe that the way they take measures is appropriate to their espoused goals and teaching methods. However, with respect to other issues such as inter-rater agreement and later-life relevance, the traditonal method has problems.

On the basis of our contact with spokespersons from a wide assortment of competency-based institutions and our analysis of program documentation, we can begin to answer questions about whether the practice of student assessment in CBE programs is markedly different from that carried out by traditonal liberal arts programs. The preceding example of the essay examination characterizes quite well traditional assessment practice, which we shall soon examine further in greater detail. We wish to find out whether CBE programs, in assessing the ability of students to perform, are fundamentally different from more traditional approaches with respect to each of the four issues set forth above.

For the sake of these comparative analyses, only the 16 undergraduate CBE programs are contrasted with the 24 traditional programs from Table 1, leaving out the Antioch School of Law and Southern Illinois University School of Medicine.

This was done to eliminate confounding of the comparison with differences in assessment practices (particularly level of competence assessed) expected between graduate and undergraduate programs.

The Nature of the Competencies Measured

In many CBE programs, the emphasis on explicit performance goals was expected to be accompanied by an emphasis on innovative and nontraditional ways of measuring outcomes. These measures can include simulations in which students might perform in a task-oriented small group discussing fabricated real-life situations and comment on them, or student observations of videotaped situations of simulated problems. Presentation measures might require students to create a multi-media presentation of a given point of view or message. Students being measured on performance might actually carry out the "real-life" adult behavior for which they have been trained, whether it is making medical or nursing rounds in a hospital, giving a piano concerto, or helping a legal services client. In some situations students might even assess themselves on these measures. These methods are very different from traditional multiple-choice or essay examinations in terms of format. But are the measures generic or specific in the aspects of competence that they consider? Are the measures used in CBE institutions to measure competence fundamentally different from measures used in traditional undergraduate programs?

We first compared CBE and traditional undergraduate programs according to the type of assessment instrument used: Fixed format "respondent" measures (multiple-choice, true-false, and sentence completion tests); written, more open-ended measures (essay examinations, reports, theses); oral presentation (speeches and oral examinations); and performance measures (behavior demonstrations, simulations, actual work performance). Table 3 presents this comparison. Simply stated, the only stable difference between CBE programs and traditional programs on this issue was in the use of behavioral performance measures, as expected. What may be a bit of a surprise is the lack of any significant difference in the use of more standard assessment techniques such as the multiple-choice exam or the written essay. Competency-based programs as a group, however, tended to use a greater repertoire of different assessment techniques. Although our small sample prevents us from drawing stronger statistical conclusions, it would appear, a priori, that the greater the range of student responses permitted in assessment settings, the greater would be the potential for convergence on the measurement of competency attainment. This remains to be seen.

Are the measures themselves more generic or specific? In the sense that the final measure of a student's performance is either largely related to a generic or a specific competency,

86

nearly all the measures used by both CBE and traditional programs are largely specific. The emphasis of many CBE program goals on explicteness is often correlated with highly specific criteria for performance, i.e., rooted in the demonstration of a specific knowledge or skill. The problem of equivalence has not been solved by our sample of programs. The tests of "critical thinking," for example, generated for German, history, or biology classes cannot be directly compared with one another. Measures of generic competence are desirable if one wants to examine a student's ability to think critically in a variety of contexts or to ascertain the likelihood that the student will demonstrate this capability after graduation. Measures of competency are not intrinsically more generic than are measures of any other human qualities, and the pressure for explicitness and detail may actually work against the design of generic measures. To be sure, those CBE institutions in our survey, most notably Alverno College and the College for Human Services, have been working toward establishing simulations and other behavioral measures that incorporate a separable generic assessment component (e.g., social interaction), and Alverno is also using measures of such generic attributes as maturity and reasoning to trace students' progress through the program. But by and large, cross-institutional comparisons of generic competency attainment have not been attempted or demonstrated.

TABLE 3

Variety of Assessment Techniques

	CBE Programs N = 16	Traditional Programs N = 24	Significant Difference
Multiple-choice paper-and-pencil	11 (69%)	21 (88%)	no
Written essays and reports	16 (100%)	24 (100%)	no
Oral examinations or presentations	11 (69%)	17 (71%)	no
Behavioral demonstrations or simulations	15 (94%)	9 (38%)	yes ($p < .01$)

The Explicitness of the Assessment Criteria

Explicitness of assessment criteria has two important aspects, as previously noted: (1) explicitness of faculty expectations of performance prior to assessment; and (2) the degree to which assessment criteria are specified and defined for assessors. In our survey, we wanted to know, for both CBE and traditional programs, whether performance expectations were mentioned explicitly or whether they had to be surmised by students; whether they were mentioned in broad, general terms or whether they were specified in considerable detail. Table 4 displays our findings. While the traditional approach to performance expectations leaves them largely unspecified, CBE programs are noted for raising faculty consciousness about explicitness of expectation. Twelve of the 16 competency programs were noteworthy for the degree to which detailed performance standards were communicated to students at least part of the time, contrasted with only 2 of the 24 traditional programs, a finding which is significant at high statistical levels.

The other important aspect of explicitness has to do with how clearly assessment criteria are specified for and shared by the assessors. This directly affects the latitude of faculty to determine privately (i.e., left to unarticulated "professional judgment") how student competence should be evaluated. As can be seen in Table 5, the majority of CBE programs have

TABLE 4

Precision of Performance Expectations by Assessors

		CBE Programs	Traditional Programs
PERFORMANCE EXPECTATIONS	Broad or inexplicit	4 (25%)	22 (92%)
	Both broad and detailed	12 (75%)	2 (8%)
	TOTAL	16	24

assessment criteria which are at least specified in broad terms for faculty members, a relatively rare occurrence among traditional programs. As we understand it, criterion specificity is more than an evaluation of "pass/fail" or "meets competency criterion/does not meet competency criterion." We gave a CBE program credit for high specificity only if categorized standards used to make evaluations were explicitly spelled out, broken down, and rigorously defined, and if there were some systematic procedures for training evaluators and monitoring their agreement. By this standard, only 25% of the CBE programs in our survey could be characterized as having approached this ideal. An institution's calling itself "competency-based" is therefore no guarantee that assessment criteria are explicit or uniformly applied, regardless of the degree to which student performance expectations are expressed by individual faculty.

Thus far, we note a certain uniformity among traditional college programs in contrast to the apparent diversity of CBE programs. Traditional programs tend to focus on generic competence, vaguely articulated, and on assessment criteria and performance expectations which are either intuitive or left to be surmised by students. Are generic competency goals and inexplicit assessment inexorably intertwined, as this observation suggests? If we consider CBE programs apart from traditional programs, we find that there is no correlation between the nature of the competency goals of the institution and the explicitness of

TABLE 5

Specificity of Assessment Criteria

		CBE Programs	Traditional Programs
CRITERION SPECIFICITY	Privately determined by assessors	4 (25%)	19 (79%)
	Loosely specified to some degree for assessors	8 (50%)	4 (17%)
	Specified in detail for all assessors	4 (25%)	1 (4%)
	TOTAL	16	24

competency assessment criteria. The articulation of goals around specific (as opposed to generic) competencies does not guarantee that assessment criteria are well defined and shared by assessors. By the same token, having generic competencies as an institutional goal does not preclude the fine definition and common understanding of assessment standards by both faculty and students (as exemplified by Alverno's program). It may be more difficult for programs with generic competency goals to develop explicit criteria, and it is certainly very costly to do so*, but they are not less likely to accomplish this task than are programs whose focus is on specific levels of competence.

The Measurement of Competency Development

With regard to this issue, our survey turned up little useful information. It is generally assumed that competencies defined as knowledge and skill change during the course of instruction, and that the measure of change is essentially the final grade or evaluation. In actuality, the final evaluation is almost always a measure of absolute attainment, not change. The conduct of assessment in both traditional and CBE programs suggests that little importance is placed on measuring what students bring into the classroom by way of competence. If measures were taken systematically before and after instruction on practice of the relevant competencies, then change in competency might be demonstrated via the measures. Merely calling a measure "competency based" does not, *ipso facto*, guarantee that it will measure change as a result of education any better than more traditional measures.

---

*FIPSE grants to Alverno College total $596,000 thus far, a great proportion of which has gone toward the assessment process.

This conclusion also holds for measures of presumably more complex or generic competencies; in particular, measures such as simulated work performance and other behavioral demonstrations which are found to be more prevalent in CBE programs. Just because a competency-based measure "looks more relevant" to the competency being assessed than does a traditional measure, that does not mean that it is a better measure of either attainment *or* change. A number of CBE programs, using simulations and demonstrations, do chart students' progress throughout the course of instruction, so change can be measured. It is not always clear, though, whether the change is due to the potency of the curriculum or to repeated exposure to the measures themselves. Neither can we examine the degree to which the capabilities that certain students have brought into the classroom relate to growth in performance on these measures, independent of curriculum content. Sensitivity to change of the measures used in CBE programs, like those of traditional liberal arts programs, is assumed but not known.

The Validity of the Competency Measure

As noted earlier, validity is a complex issue. In reviewing program descriptions with other documents from CBE programs and our sample of traditional programs, we sought to understand differences among institutions regarding both rational validity (including face and content validity) and statistical validity (including construct, concurrent, and predictive validity). The results of our conversations and reading are presented for each of the relevant validity issues.

Face validity

Regardless of whether a measure taps anything of value, it must at least "look like" it does. Indeed, nearly all the traditional programs surveyed and 13 of the 16 undergraduate CBE programs presently recognize no other kind of validity.

Given the politics of higher education, face validity is necessary for any measure that is going to be taken seriously, even if on psychometric grounds the measure is problematical. Professors simply will not consent to evaluation of their teaching or of their students' learning by measures that, in their view, trivialize the learning process or do not obviously represent that which is to be learned. One of the effects of a college education, according to Feldman and Newcomb (1973), is to make students more theoretical and perhaps less religious. A research psychologist might try to measure the effect of college on students by assessing this presumed change using test "items," such as the following one adapted from one of the most widely-used questionnaires in educational research, the Allport-Vernon-Lindzey Study of Values.

Which of the following two headlines would you be more likely to notice and want to read more about:

(Check one)

(a) SCIENTIST ANNOUNCES NEW THEORY (theoretical)

(b) THEOLOGIAN DISCUSSES NEW DOCTRINE (religious)

After college, a student should tend to change from checking answer "b" to checking answer "a". "But," the professor protests, "that isn't what college is about at all." The complaint might run: "Some psychologist has used a meat cleaver where a scalpel was required." Most professors simply will not consent to have their careers and reputations rest on check-mark answers to questions like that, and most students would strenuously resist the use of such an item to determine if college classes have had an impact on them. Parents, employers, an maybe even the courts would also take a dim view of that kind of measure as the basis for students' grades, degrees, or credentials, or teacher promotion and tenure.

Traditional psychological tests, not surprisingly, are seldom used to measure the effects of higher education. Isolated programs may include the administration of aptitude or personality measures as part of an institutional research program, but as serious measures of the effects of higher education, measures that could be used to credential student competence attainment and to evaluate programs, they simply will not work because faculty do not believe in them. Most faculty insist that measures of effects look like they measure what they claim to measure, and that they measure outcomes at the same apparent levels of complexity, sophistication and subtlety as those which characterize the relevant processes of instruction.

Content Validity

Content validation is a rigorous procedure by which a test is designed to represent the domain of knowledge, skills, traits, and other qualities relating to task performances that are included in instructional objectives. Though we were unable to get clearcut evidence of content validation within the samples of traditional programs, we may safely conclude that formal content validation is extremely rare here, that it is an assumed property of all faculty-designed measures of student performance. As far as CBE programs are concerned, we find little evidence to suggest that the average state of affairs is any different. Only four of the 16 undergraduate CBE programs surveyed (Alverno; The College for Human Services; the University of Massachusetts, College III; and Mt. Hood) have taken steps to ensure that assessments adequately sample the knowledge and skill domains of curriculum content. Content validation of assessments is unusual primarily because this aspect of validity is also viewed as self-evident (and therefore often confused with face validity in the minds of those with whom we spoke), and only secondarily because of the costs involved.

A practice which may be construed as yielding some semblance of content validation and thereby legitimizing the assessment process is that of employing assessors external to the institution to evaluate student competence. This is believed to be of particular value if the external assessors are members of the same professions for which students are being educated, for if the assessors participate along with faculty members in the process, an implicit stamp of approval is given and relevance to competent performance is assumed. As Table 6 illustrates, CBE programs are much more likely to use external assessors (including practitioners and supervisors on jobs) than traditional programs, though nearly 30% of the traditional programs surveyed do use outsiders for assessment purposes.

But, in fact, is the use of external assessors more valuable than simply restricting the function to faculty? To this question we could get no satisfactory answer. Of course the opinions of expert panels and job supervisors would be expected to carry a certain weight over and above faculty judgments, but there are no data to suggest that these expert judgments are any more reliable or valid. Recent research by Kane and Lawler (1979) suggests, in fact, that the judgments of performance by experts (i.e., job supervisors) is less reliable or valid than judgments of performance by students (i.e., one's peers). Interestingly, especially in light of these studies, student peer evaluations were used in 7 of the 16 undergraduate

TABLE 6

Assessors of Student Competence

		CBE Programs	Traditional Programs
WHO ASSESSES COMPETENCE	Faculty only	2 (12%)	17 (71%)
	Faculty, with outside assistance	14 (88%)	7 (29%)
	TOTAL	16	24

$x^2 = 10.86$

CBE programs, but were not used in any of the 24 traditional programs in our sample. As far as can be determined, however, peer assessment does not carry very much weight in any of the CBE programs in comparison with faculty and outside expert judgments. In sum, legitimacy appears to be the key benefit of using outside assessors. With that consideration aside, there appears to be little significant difference between CBE and traditional programs insofar as content validation of assessment is concerned.

Construct validity

Construct validation is a highly technical process which is best carried out at the institutional level, since it involves statistical comparisons among multiple measures. Under separate grants from FIPSE, McBer and Company and the American College Testing (ACT) Program have undertaken this kind of validation for their own tests, involving overlapping sets of colleges, including a number of CBE programs as a base. Construct validation conducted intramurally on competency measures used in the classroom is difficult for most colleges to bear without research funding. In fact, only Alverno College, among the undergraduate CBE programs in our survey, is in the process of collecting this information with the assistance of NIE funding for this purpose. Only one other CBE program had attempted construct validation as part of institutional research, but the project was abandoned when funding ran out.

Construct validity, like face and content validity, however, does not bring us any closer to an understanding of a measure's relevance to performance in later life. It does tell us more about what we are actually measuring but little about the value of what is being measured, except insofar as a test is correlated with a separate measure that is related to important life outcomes.

Concurrent and Predictive validity

As described earlier, concurrent and predictive validity are different forms of criterion-related validity. Concurrent validity is expressed by the relationship between a measure of a personal characteristic and some criterion at a given point in time (for example, between age and salary, or between verbal *ability* and college grades). Predictive validity describes the degree to which a personal characteristic measured at one point in time relates to some criterion determined at a later point in time (for example, how socioeconomic status at birth relates to adult salary, or how verbal *aptitude* measured before college admittance relates to college grades).

In either case, the choice of a criterion is a critical issue. The worth of a competency-based assessment procedure is ultimately no greater than the worth of the criterion measure to which it is related. College grades are popular criterion measures for validation purposes, under the assumption that college grades themselves relate to something of importance other than "educability" in classroom situations. But as Heath (1977) has noted, grades are poor predictors of successful life adjustment and career performance. The consequence is that, for CBE to fulfill its promise, measures must be related directly to important work and life outcomes.

Our survey of CBE programs, however, found this promise to be unfulfilled as far as the demonstrated relevance of competency measures to performance in later life is concerned. Not a single program in the sample, either competency-based or traditional, had actually completed the steps needed to show the concurrent or predictive validity of its measures. To be fair, though, we must recognize that predictive validation is a long-term process, since students must be tracked a number of years out of the program in order for us to conclude that assessments in college were related to later life performance. None of the CBE programs in our sample have been in existence long enough to be offered that luxury. Concurrent validation, on the other hand, is another matter, for it requires the identification of people at varying levels of competence who are already part of the workforce and the demonstration that more competent individuals perform better on the assessment instruments than their less competent counterparts. Nevertheless, none of the 16 undergraduate institutions in our CBE sample have documented such concurrent relationships to date for their assessment measures.

To pursue this issue further, we asked if validation studies were even being considered by the CBE programs in our sample, not making the distinctions among content, construct, concurrent, or predictive validation. In doing so we found that seven of the sixteen CBE institutions were indeed either considering a validity study or were in the process of conducting one. Disappointingly, however, except for Alverno College and the College for Human Services, most of the CBE Programs were not planning either concurrent or predictive validity studies.

The later-life assessments in the great majority of CBE programs thus seems to be more a matter faith than demonstrated evidence. A measure cannot be assumed to be valid if it is called "competency-based" or even if it "looks more relevant" than traditional measures do. A professor may design the most realistic simulations of real-life, on-the-job performance and yet, without further evidence, the conclusion that this measure has real-life relevance involves the same leap of faith that a more traditional professor would make about the Freud vs. Maslow essay question. No matter how realistic, a simulation is a simulation and not a real-life performance, and college is college and not a career. A variety of influences including ecological factors (setting, presence of evaluators or TV camera, etc.), individual response sets (knowledge of evaluation), and personal dispositions (such as shyness, test anxiety, and evaluation apprehension), ensure that an evaluative assessment is never going to be identical to "real life." (For example, presumably "real" medical rounds made with an

103

evaluator, in the knowledge that the evaluator is monitoring performance, are not the same as the actual rounds of a practicing doctor.) Of course, it seems reasonable to believe that these simulations are pretty close to real life, or that they are at least closer than are traditional methods of measurement, especially, for example, if students say that the new methods "feel more relevant." Such beliefs are, however, just that: beliefs. Of course, evidence can be gathered for any measure, traditional or nontraditional, explicitly competency-based or not, to demonstrate whether it is relevant to real-life careers and roles. But in the absence of evidence suggesting that the measures correlate with performances in real-life settings, even the measure that feels or seems most relevant rests on unsubstantiated faith which may lead professors in either the CBE or traditional programs to erroneous assurances about later-life performance.

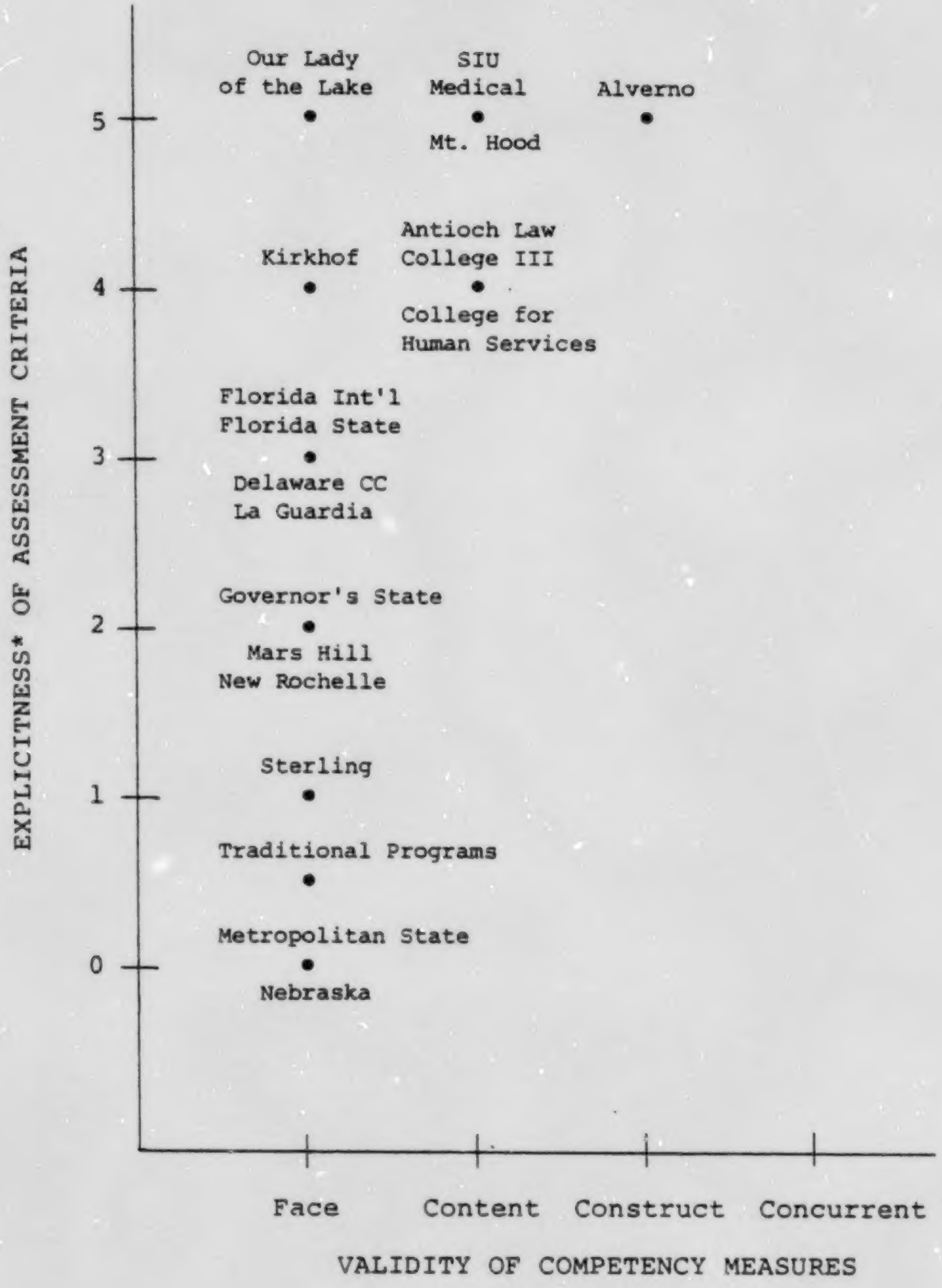
Cross-Classifying Programs by the Two Most Critical Issues

In our judgment, the two most critical issues of those discussed above are the explicitness of the assessment criteria and the relevance of assessment to later-life performance. The nature of the competencies (i.e., whether they are generic or specific) and the measurement of competency development are issues that are no less pertinent for traditional programs than for CBE programs. The levels of competence addressed derive from institutional goals, and competency development should be a concern of both CBE and traditional programs, but these issues are not ones on which the CBE movement rests its claims of uniqueness. The original intent of reformers in educaton, that CBE programs be "student centered" and "outcome-driven," suggest that these programs be measured against standards of explicitness and relevance or, in other words, that assessment criteria be clear to students and shared by assessors, and that the competencies measured be demonstrably linked to adult performance.

Figure 2 brings together the discussion of these two issues and classifies the 16 undergraduate CBE programs, the two additional CBE programs at the graduate level referenced in Figure 1, and the grouped traditional programs. Programs are arranged according to the explicitness of the assessment criteria (combined estimate of the clarity of student performance expectations by faculty and the specificity of the assessment criteria for assessors) and the quality of the validation efforts conducted to date (face, content, construct, and concurrent/predictive validity, in order of increasing importance to the goals of CBE). What we find is that, compared with traditional undergraduate programs, the majority of CBE programs are more explicit in one or more aspects of their assessment criteria. Students in these programs have more information available to them about faculty expectations and there is a greater sharing of assessment criteria among faculty and other assessors.

FIGURE 2

CBE and Traditional Programs



Nevertheless, we also observe that, for all the traditional programs and most of the CBE programs surveyed, face validity is the prime test of a measure's relevance to later-life performance. The remainder of CBE programs have performed content and construct validation for some of their measures, but to date none have been able to demonstrate relationships between measures of competency and later-life performance goals. The likelihood that an institution will undertake validation of any sort beyond face evaluation of measures, however, appears directly related to the explicit uses of the assessment criteria. This seems both logical and sensible, since anything beyond face validation requires that institutional goals be stated in ways that allow them to be tested, and the specificity with which competencies are described enables such scrutiny.

The pattern of results observed leads us to a conclusion that can be applied to the majority of programs that call themselves "competency-based:" Although the language of CBE is replete with references to the "relevance" of both the competencies and the methods used to assess them, this relevance rests on a priori assumptions and not on any demonstration of relevance. In terms of epistemology and the standards of evidence, most CBE programs are indistinguishable from traditional undergraduate programs. The major differences between CBE and other programs is the explicitness with which competencies are described. To be sure, there are other differences in approach to assessment, including more use of behavioral assessments and outside assessors, but no evidence is yet forthcoming as to whether nontraditional behavioral measures are any more related to later performance than traditional modes of measurement, or whether external evaluators provide assessment data of greater validity than do their faculty counterparts.

The change of programs to being designated as competency-based seems, in the main, to be more one of process than of outcome, more one of specificity than of relevance. Perhaps it is inevitable that this should be so. Since most educators are not practitioners as well, we might expect that their commitment to rational, as opposed to empirical analysis of what they are doing, and their preference for theoretical, philosophical, and ideological matters will always prompt them to design curricula and student evaluation on the basis of their own discussion and analysis rather than on the basis of what can be demonstrated empirically. While in some cases it is possible to design curricula and test for learning on the basis of empirical relevance, or to plan the test of _a priori_ assumptions about competence with comprehensive, methodical analysis at the program level, these tasks may be unreasonable for most educators to undertake. While CBE may have roots in American science and pragmatism, as well as urgent social necessity (Grant, et al., 1979), as an organizational movement it may be more of a trend in language and fashion than a shift in paradigm from the speculative to the scientific.

V. AN ALTERNATIVE APPROACH TO CBE

Competency-based education, as it is carried out in practice, is very much a compromise between new goals and old values. Pressures on institutions from the workplace, declining enrollments, and national priorities for more "relevant" education run into resistance from the academic community. This is especially true in the area of undergraduate liberal arts education, and to a lesser extent in other strains of CBE. The reasons for this state of affairs are persuasive, including the well-established philosophic tradition underlying liberal education, perceived threats to academic freedom arising from substituting empiricism for independent ideology, and a reluctance to accept the abilities required in the world of work as defining the domain of educational goals. The overall aims of CBE programs are not as much at issue here, as are the processes by which those aims are achieved.

One of the major reasons that CBE has yet to demonstrate its potential as a movement is that the current CBE programs were established in institutions that were already in existence. Changing a program from traditional to competency-based is a major organizational intervention. Only where the intervention has been undertaken institutionwide, and not without considerable stress has progress been made toward demonstrating the life-relatedness of the institutions' curriculum goals. CBE programs that are introduced in traditional institutions have built-in obstacles to overcome. When program change is not an economic necessity, aren't broad institutional goals, such as enabling a "broad perspective" and a "fuller life," justified in the reported satisfaction of alumni, perhaps expressed in contributions to their alma mater? And regardless of an institution's financial state, isn't face validity the accepted standard against which to judge a program's worth? We do not argue that college cannot implement CBE programs which measurably improve the quality of graduates in terms of their performance in the workplace, just that with the exception of a handful of CBE programs, there is little internal incentive to do so.

Most of the programs we surveyed have begun with a priori notions of what competence is, have designed curriculum and assessment measures, and then have attempted or are attempting to draw (rationally, seldom empirically) a relationship between curriculum objectives and career performance. Rare is the program that has had the luxury of beginning with a clear understanding of educable factors that make real differences in life roles, and then working toward program development and evaluation. But such programs do, in fact, exist, largely outside of mainstream secondary and postsecondary education. Three examples of programs which have developed or are developing along these lines are the following:

1. The Univeristy of Minnesota College of Pharmacy. As described by Cyrs (1978), development of this competency-based pharmacy program grew out of a theoretical definition of competencies as composites of many performance objectives determined by situational outcomes, rather than as abilities to perform discrete tasks. An important addition to this definition was that competencies must be data-based and related to important work outcomes. To ensure that this was so, a task analysis was performed on the job of pharmacist, resulting in 34 competencies. But this study went one step better: It collected over 800 "critical incidents" pertaining to pharmacy practice from 34 faculty members and practitioners, using methods advocated by Flanagan (1954) for interview data gathering. Analysis of the critical incidents yielded 9 performance dimensions derived from the 34 competencies from the job analysis, plus 11 personal trait dimensions important to performance (e.g., decision-making, independence, behavioral maturity) which did not derive from the task analysis. The pharmacy curriculum and assessment practices were developed with these performance dimensions explicitly in mind, and both are being evaluated against an "externship" experience designed to replicate the conditions of practice. The assessments, in particular, are quite varied, including paper-and-pencil tests, free response essays, individual simulations and behavioral demonstrations, and a complete assessment center.

2. The Navy Leadership and Management Education and Training (LMET) Program. The research for this program began in 1976 when more than 1,000 Navy officers in all levels of service were studied to determine competencies which distinguished superior from average leaders and managers (Winter, 1979). From this research emerged a series of two-week courses through which all Navy officers will pass at different stages in their careers. These courses teach specific job-related management skills, but the aim of the program is to impart generic clusters of competencies, which underlie concern for efficiency and effectiveness, skillful use of influence, conceptual ability and problem solving, process management, and command control. Assessment and program development have gone hand in hand, and designed into the program is a plan to evaluate the curriculum by monitoring the performance of its graduates.

3. The American Management Association Masters in Management Program. The basis for this program was a reanalysis of more than 13 independent studies carried out on different groups of managers by McBer and Company over a 10-year period (Boyatzis, 1980). Managers at all organizational levels in a number of job functions in the private and public sector were represented, including one sample from a longitudinal study of managerial effectiveness. In the analysis of competencies, outstanding managers were compared against less

effective managers as determined by a number of criteria, including:

- "Hard" measures which are directly related to objective performance goals (e.g., sales, profitability, and productivity);
- "Soft" measures which are largely attributional (e.g., supervisor, peer, and subordinate performance ratings or nominations); and
- "Indirect" measures having an assumed but not necessarily valid link to hard outcome measures of job performance (e.g., speed of promotion, salary, turnover, and subordinate morale).

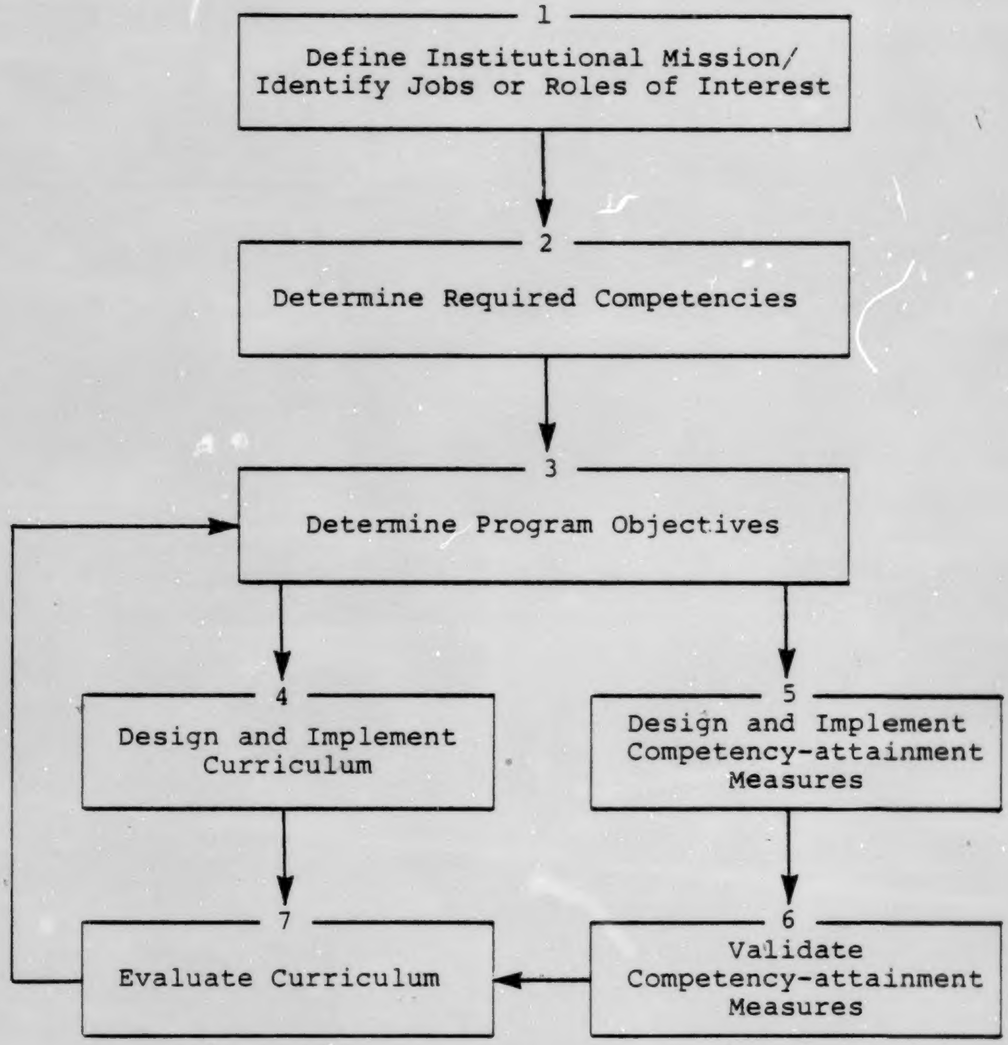
Five competency clusters guide the development of this course: entrepreneurial, intellectual, socio-emotional maturity, interpersonal, and specialized knowledge. The program that is being developed is a one-year course for nontraditional students: managers who have had previous job experience. All students are to be assessed at entry, given credit for important competencies they already possess, and charged to develop competencies which they lack. Assessment at entry, during the program, and at exit will determine both the efficiency of the program and the readiness of the student to be a more competent manager.

These empirical approaches to competency identification were developed out of a need to identify personal qualities which relate to concrete organizational goals, with the realization that the degree to which the educational process is responsive to those goals will dictate the degree to which the programs succeed or fail. This set of conditions is absent from the majority of educational institutions whose faculty have no formal ties to the careers for which students are being prepared, and this partially explains the lack of pressure to validate the job-relatedness of curricula in these institutions. One might counter, however, with the argument that the program goals of organizations using empirically-based models are more limited in scope, and the competencies identified and assessed are therefore more trivial than those of their academic counterparts in the CBE movement. The answer to this challenge is that many of the goals of the former group of organizations, including effective leadership, service, management, and administration, involve competencies similar to those included in the mission statements of the majority of the CBE programs surveyed.

What is most important, for the purpose of this report, is that each of the empirically-based programs cited shares a development process. The major components of this process were explicated by Goldstein (1974) in his book on training design and evaluation, and incorporate the work of Cyrs (1978), Winter (1979) and Boyatzis (1980). The steps in the process are illustrated in Figure 3, and in greater detail in the following pages. It is the goal of this section to build a case for a viable approach to implementation of CBE and then to examine the extent to which CBE can benefit from including parts of this process. As will be seen, various components already exist in a number of current CBE programs.

FIGURE 3

Competency-based Assessment and Program Development Process


1
Define Institutional Mission/ Identify Jobs or Roles of Interest
2
Determine Required Competencies
3
Determine Program Objectives
4
Design and Implement Curriculum
5
Design and Implement Competency-attainment Measures
6
Validate Competency-attainment Measures
7
Evaluate Curriculum

Steps in the Process

Step 1. Define institutional mission/identify jobs or roles of interest.

An educational program which aspires to being competency-based must begin with an attempt to state its long-term institutional goals or mission. This holds whether the program be a four-year liberal arts experience or a four-week course on management skills. Such a statement of purpose is not the same as program objectives for what a student would be expected to learn in a series of courses, for it is premature to speak of these shorter-term goals without an understanding of what competencies are important to impart. Identifying the jobs or roles of interest must proceed in parallel with defining the institutional mission and on the competencies that are taught, for each has an effect on the other. For example, where the nature of the role is diffuse, goals tend to be broadly stated, implying a focus on generic competence, and this is true of most liberal arts programs. Where the nature of the role is more specific, as a particular job would be, the institutional mission may be phrased in language that is job-focused, implying a concern with competencies that are more specific than generic. Restraint must be exercised at this point, lest the particular competency goals be driven by the statement of mission prior to a determination of which competencies are important.

This initiating step can be carried out in a number of ways. Organizations outside of the education mainstream may approach the mission identification process by first determining the kinds of jobs or functions that are presently required, and then devote resources to either selecting or educating individuals to carry out jobs or fulfill the functions. Another approach may be to project future capability needs, based on current trends and knowledge about the characteristics of the workforce and on anticipated challenges and new conditions that will affect the organization. In this way, educators can begin to identify jobs and roles that will play a prominent part in the future life of an organization or society but for which there are relatively few capable incumbents.

Finally, educational programs may be forced to respond to abrupt changes in political or economic conditions with institutional goals which implicate a different mix of jobs or roles than those previously valued. Regardless of how mission identification is carried out, the institution should still attempt to isolate the jobs or roles but not adopt a final program plan without further examining the qualities of the individual that are demonstrably related to effective performance.

Step 2. Determine Required Competencies.

This step is really a whole process in itself which can be broken down into a set of two more manageable components:

identifying the competent individual and analyzing that individual's performance.

(1) Identifying the competent individual. The competent individual is primarily identified in the outcome he or she obtains, not in the means by which he or she obtains those outcomes. This is to say that any ability, skillfully employed that does not yield effective results in a given job or role is not a competency in that job or role. So, rather than asking whether certain characteristics are related to good leadership, citizenship, or life adjustment, for example, we should first ask what are the outcomes that indicate effective performance in these roles.

Each job or job family has its own implicit or explicit criteria for effective performance. Sales and production jobs are fairly easy to assess because the indicators (i.e., sales volume, error rate, or efficiency) are generally available. Even human service work can be assessed on the basis of avaiable outcome criteria: In a study of Navy alcoholism (Boyatzis & Burruss, 1977), performer effectiveness was documented by tracking the number of individuals with persistent alcohol problems who remained on the job six months after treatment. Management jobs, by contrast, may require indirect assessment, for success in these is due mainly to the effective performance of subordinates. In this case, measures of subordinate production and efficiency or survey-based measures of morale and attitude may provide an effective, if indirect, indication of supervisor performance. Additional, "soft" data such as peer and supervisor ratings are available for most jobs. Nomination by peers is a reliable source of data that can identify the good performer in the absence of stricter, more directly observable outcome measures (Kane & Lawler, 1979).

When "life effectiveness" is chosen as a criterion, however, there is bound to be heated debate about what outcomes constitute "competent performance." In the past, researchers have posed such measures of life effectiveness as salary level, marital adjustment, mental and physical wellbeing, frequency of job change (stability), life expectancy, community involvement, self-sufficiency and adaptation to new conditions and circumstances. In part, the goals of the institution will decide which of these criteria, if any, are relevant indicators of effectiveness.

(2) Analyzing the competent individual's performance. Once individuals have been identified who satisfy one or more criteria for effective performance, the next step in a competency development plan is to find out what competencies these people possess. One approach to identify competencies that has been used in the workplace involves the "expert panel," in which experts on the job to be analyzed and/or job incumbents are asked to suggest areas of knowledge, skill, or behavior that appear characterisitc of the competent performer. The qualities, skills and abilities determined through this method,

however, are likely to include a combination of specific behaviors (e.g., "adheres to cost accounting procedures") and vague personal qualities (e.g., "integrity" or "critical thinking"). The product of this approach is only a collective theory of what constitutes competence, a theory which must be validated in some other way before an institution seriously can undertake designing its program.

Another approach to competency identification is job/task analysis. If task analyses are to be undertaken, one should compare the performances of outstanding career incumbents in contrast to their average counterparts to see if they approach tasks differently. In certain jobs it is practical and realistic to expect to observe differences in behavior which signify qualitative differences in overall performance. However, in the majority of complex jobs and careers, it is not efficient to catalog all the behaviors that comprise performance, since most of what is observed is routine behavior unrelated to performance in critical situations. Also, task analysis procedures typically yield minimal competencies (threshold skills) rather than optimal competencies (related to outstanding performance). The University of Minnesota College of Pharmacy program, for example, used a combination of panels and task analyses to identify competencies (Cyrs, 1978). Task analysis produced a list of competencies related to adequate threshold performance, while the panels provided more specific statements about personal qualities hypothesized to relate to good performance as a pharmacist.

An alternative to task analysis for identifying competencies is the thematic-analytic approach taken in the Navy LMET and the American Management Associations programs. Rather than defining competencies as behavioral objectives (c.f., Mager, 1975) or as the performance of discrete tasks, the thematic-analytic method takes the position that competencies are human characteristics that underlie effective performance, rather than the observable behaviors themselves (Klemp, 1979). This method contrasts the actions, reactions, thoughts, and attitudes of competent or outstanding performers with those of less effective performers in the same job to discern competency themes that organize behavior and thoughts in a variety of situations. The thematic-analytic method, an inferential process, is designed to identify competenices which are generic and optimal in nature. But like any other means of competency identification, it is not self-validating.

Step 3. Determine Program Objectives.

With the statement of institutional mission and the preliminary identification of competencies related to the jobs or roles of interest having occurred, the institution can proceed to define program objectives in greater detail. A competency "model" of the effective adult performer should serve as the framework for developing these objectives. A competency model is nothing more than a coherent statement of competency goals for the curriculum, stated to represent the ways in which competencies contribute to achieving effective outcomes, or how underlying performer characteristics enable individuals to meet performance criteria. Some competencies in the model may be fairly specific, task-related and mundane, and though they may not distinguish the outstanding from the average performer, they are required by both. Thus, the program objectives should make a clear distinction between these "minimal" competencies and other, "optimal" competencies which do contribute to outstanding performance.

In addition, the fact that some competencies are more difficult to develop than others must be taken into account. For example, self-confidence may be a personal characteristic related to effectiveness in a job or role, but as it is much more difficult to change in students than their knowledge base or skill repertoire, it needs to be treated differently in the program objectives. A program must make difficult decisions about which competencies it will broaden and which it will ignore, given constraints on time and resources. In some programs, it may be wise to select students on the basis of competencies on which the program can have little effect but which nevertheless are crucial to effective performance, much as the scholastic aptitude test is currently used to determine admission standards. In other cases where it is safe to do so, a program may ignore the minimal competencies and concentrate its energies on developing the optional competencies (or vice versa).

In any event, the program objectives should be stated in language that is specific and precise, rather than general or vague. To enable the eventual validation of both curriculum content and competency measures the objectives must also be stated in such a way that allows them to be measured against the same criteria by which the competencies were derived; that is, successful graduates of the program should achieve the same outcomes as the effective performers whose competencies are represented by the curriculum. We have already seen in our sample of CBE programs that the more specific the program objectives are in ways that lend themselves to measurement, the more likely the program is to attempt something beyond face validation to demonstrate its relatedness to desired perfomance outcomes.

Step 4. Design and Implement Curriculum.

Designing a competency-based curriculum for adult learners is in many ways a different task than developing course materials for traditional students, at least in theory (Knowles, 1970). In practice, however, adult learners of all ages do have a number of things in common. They all bring to the classroom previous life experiences and some idea of what they are being educated to do: They are not passive receptacles to be filled with knowledge, nor are they without concern for a curriculum's relevance to life after graduation. Competency acquisition is an interactive process that can organize an educational experience, independent of a particular area of specialization that a student may choose.

The statements of institutional and program goals should therefore provide students, at the outset, with a clear understanding of the competencies or qualities the educational institution is committed to imparting. There should be opportunities in each program of instruction for students: (1) to recognize the relationship between that which is being taught and competence in adult life; (2) to understand these relationships in terms of their own ways of thinking and based on their own experiences; (3) to assess themselves against specific competencies in terms of skills applied in working with the subject matter for a particular course; (4) to develop competencies (in the context of a program of instruction) which have been identified specially for improvement; and (5) to apply the knowledge and processes learned to experiences outside the classroom. These principles are the basis of successful programs in industry and are neither new nor unique to CBE. Where they have been applied the most is in training programs designed to impart a more limited array of basic skills and competencies. However, the basic ideas are equally applicable to liberal arts curricula: in order to facilitate competency development, students must value the competency, understand how it relates to competent performance, and be able to test and benchmark their ability to perform competently in different situations.

For pedagogical and psychometric reasons, curriculum development and design of assessment techniques proceed along different lines. Nevertheless, some of the issues discussed earlier surrounding the practice of assessment in contemporary CBE programs also apply to curriculum:

1. *Re* the nature of the competencies, provide learning situations that allow repeated practice of specific skills, but also provide a variety of learning opportunities for generic competencies so that students can generalize their capabilities.

2. *Re* the explicitness of the competencies, make certain that the general principles underlying the curriculum are clearly understood and maintain consistency between coursework and program goals.

3. *Re* the developability of the competencies, track student performance at key intervals in the competency development process either formally or informally, and share these findings with the students themselves.

4. *Re* the validity of the competencies, maximize the apparent similarity between what is to be learned and the outcome performance of interest (i.e., establish face validity), but also develop an explicit rationale for the competencies emphasized in the curriculum and look for curriculum objectives that reflect desired outcome performance in later life or job roles.

Step 5. Design and Implement Measures.

Although development of curriculum and assessment should proceed concurrently, one should not be designed with the other in mind except as they focus on the same performance criteria as stated in the program objectives. It is common practice to assess learning of the curriculum content rather than the criterion performance of interest, but this is not sufficient here.

Criterion referencing is the heart of the competency-based approach to assessment. Messick (1975) contrasted criterion-referencing with norm-referencing in this way:

> "A norm-referenced test is one that is constructed to yield test scores that discriminate among individuals on the trait measured by the test and that are interpretable in terms of the relative performance of other individuals and groups on the same test. A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specific performance standards. At the level of interpretation, the distinction seems clear: A norm-referenced interpretation compares an individual's test performance with the performance of others; whereas, a criterion-referenced interpretation compares it with a performance standard."

Norm-referenced measures may tell us how students compare to each other in the classroom, but they tell us little about what absolute performance levels are demonstrated with regard to standards of competent performance.

If measures of competencies can be developed to distinguish effective from less effective performers, the performance of the former group of individuals on these measures may serve as the standards to which students would aspire. Students of course still might be graded relative to one another, but they should know where they stand relative to competency levels. A competency measurement system should have the following characteristics:

(1) It should provide the educational institution with information about the proportion of its students who are performing well, relative to the competencies established as program goals.

(2) It should inform students of how well they are performing relative to the competencies desired as well as relative to each other and to individuals in other programs.

(3) It should not be susceptible to observer biases related to personal characteristics such as age or race, or other qualities that are irrelevant to the demonstration of the desired competencies.

(4) Finally, it should respond to the important issues relating to competency-based assessment: The measures should assess generic and specific competencies, the assessment criteria should be clearly specified for students and assessors, the competencies measured should be sensitive to program-induced change and the measures should predict later-life performance outcomes.

Step 6. Validate Competency Attainment Measures. Measures used to track competency attainment may have two purposes:

to provide an accurate reflection of an individual's competency attainment; and to measure the effect of the program on students in general. Face validity and explicit (i.e., shared) assessment criteria are not necessary properties for a measure to be a useful indicator of overall program effectiveness, but are preferred for classroom administration. Regardless of the use to which an assessment device is put, the measures should be analyzed to determine their validity.*

The primary determination of the worth of a measure is based on whether it measures anything of meaning. For the purposes of outcome-oriented education, this implies that a test should be positively correlated with effective performance against the criteria specified in the program objectives. At the very least, the institution should determine if effective adults (by whatever criterion in whatever role) perform better on the measure(s) than do their less effective counterparts (concurrent validation). This procedure is one example of criterion-related validation. A stronger test of a measure's worth is predictive validation, or whether a person who performs well on the measure is found later to be more effective in his or her or role than another person who performs less well on the same measure. Predictive validation strategies are, by definition, time-consuming and not very practical when one is trying to develop a curriculum and bring a whole program up to speed at the same time. Nonetheless, if an institution is able to undertake the outcome criterion identification process (Step 2), it should be able to proceed with a program of concurrent validation.

---

*Reliability of assessment measures is not addressed here. Test reliability has been a minor problem in most CBE-based assessment programs in comparison with validity. Chapter 1, devoted to employment selection, discusses in detail the vicissitudes of the reliability-validity issue.

Once a measure has been shown to "predict" performance, construct validation becomes more important. Construct validation entails a set of statistical procedures whose purpose is to determine the relatedness of individual measures to each other, and is necessary to ensure that a test measures the characteristic it is thought to measure. Many tests that are face-valid, especially nontraditional behavioral or simulation measures, may really be measuring dimensions other than those intended and need to be checked out. Educational validity is an additional aspect to be considered: Is the measure sensitive to changes in the competency which may occur during the program of instruction? This is a subtle issue, because an instrument capable of assessing change may not do so if the curriculum is ineffective. Measures of aptitude and personality characteristics were mentioned earlier as two types of instruments which are designed expressly to be stable over time and so have built-in barriers to their utility in CBE. Wherever possible, a program should adopt measures which have demonstrated their sensitivity to measuring individual change in different settings.

Step 7. Evaluate curriculum. Are students any better off after their educational experience in ways that make a difference in later life? Certainly the student will have gained a wealth of specific content knowledge through classroom experience, but is he or she subsequently able to apply that knowledge effectively in critical life and career situations? Has the student indeed become more competent? If we have been faithful to the first six steps in the program development process, we should be able to answer these questions. Specifically, if a program has developed measurements of individual competency attainment that are related to effective performance outcomes and that satisfy the important criteria for educational assessment, we can identify changes in those measures for a particular student between the beginning and the end of the formal educational experience. One study of the "value added" by liberal arts educational institutions has been carried out by Winter, McClelland, and Stewart (1980) on a set of generic competency measures. The time required to carry out studies like this depends upon the length of the program of instruction, and for a liberal arts college, four years may be too long to wait for results from a new program. One may take intermediate measurements at critical points in a student's career to gain an indication of whether particular programs have been successful at developing competencies. Ultimately, the outcomes of the new graduates compared with those of the graduates who went before will provide justification for the new approach.

In employing organizations, one can get more specific measures of effectiveness and look for measurable changes in job performance which can be directly related to the competency-based training program (assuming proper controls; i.e., that these individuals are compared with others who experienced a different program of instruction or did not participate in the new program). Whatever the form that these program evaluation studies take, they should be used to refine the other steps in the process, all the way from determining program objectives through the validation of competency attainment measures, as Figure 3 illustrates.

How the Process Relates to Assessment in Current CBE Programs

The process just presented for developing CBE programs is, in many ways, a representation of the ideal progression from defining institutional mission to program evaluation. The question remains, however, as to whether this ideal is practical. To see if this is so, we can review our knowledge of the eighteen CBE programs in our sample to see how each of the seven steps in the competency-based assessment and program development process is typically approached and the consequences thereof.

To put the comparison in perspective, we note that traditional liberal arts institutions proceed from a statement of mission to program objectives and then directly on to the curricculum development and implementation with little or no subsequent program evaluation. As we have seen, CBE programs are characteristically different in some ways from traditonal programs. The purpose of the following discussion is to examine which attributes of CBE programs can be modified to approach a realistic ideal for the assessment of competence.

Defining Institutional Mission and Program Objectives

All of the CBE programs in our study began with more or less detailed statements of institutional mission and a focus on particular jobs or life roles for which students would be prepared. But too often, the institutional mission was extended to include shorter-term program goals before much was known about the competencies required for these jobs and roles: The goals are all *a priori* goals. The founders of many traditional universities also had strong ideas about what they wanted to accomplish, including what subjects should be taught (or not taught) and which issues (moral, spiritual, mundane) would be emphasized, but with few exceptions, the question of whether such goal statements bear a demonstrated relation to adult performance has not been answered by CBE institutions.

In practice, statements of institutional mission tend to have lives of their own, regardless of their validity, and tend to drive the competency specification process at the program level. If we consider a statement of mission as reflecting an institution's self-image, we can ask whether that self-image is reflected in student development. Winter, McClelland, and

Stewart (1980), as part of their research on the effects of liberal arts education, examined the relationahip between an institution's self-description, as presented in the college catalogs, and changes in general psychological characteristics of students from freshman to senior year. The results showed that institutions which described themselves in terms of specific characteristics tended to develop those characteristics in their students.For example, one college which described itself as characteristically proactive and challenging showed significant effects in changing a student's self-concept as a capable, independent actor. Another institution, which described itself as a warm and friendly place, showed a marked increase in students' need for warm, close relationships. Some of these results may be artifacts of students' self-selection into institutions whose self-image matched their own, but while this would explain differences in initial level of these characteristics, it would not, by itself, explain the changes wrought by four years of encounter with the institutional milieu. Perhaps these changes were desired by the institutions, but if true, there were no attempts to measure such changes as part of the college assessment program. In any case, there were no data before the fact to suggest that these changes were more or less important than other program effects on a student's competence after college.

Practically speaking, there is little that can be done to modify statements of mission for existing CBE programs. What can be done, however, is to clarify the roles for which the institution is preparing its students and then ask whether the mission statement adequately serves these roles. One approach might be to survey what careers alumni are engaged in and then to ask whether the mission statement has served its intended purpose. The aim of this simple process is to enable institutions to examine some of the consequences of their programs. For example, if the mission of an institution is to develop leaders in the public service, how many graduates actually fulfill this role? And if the goal of the institutions is to promote personal adaptability and growth, how many graduates are satisfied with their lives or are employed at or near their level of aspiration? Without some check on whether an institution is attaining it goals, there can be no movement toward other changes in educational practice. Knowledge of a discrepancy between the valued institution goal and the accomplishments of graduates would create pressure for change.

Determining competency requirements. The immediate prerequisite for this step is a clear definition of institutional mission and with it the specification of jobs or roles implicated by that mission. The most difficult process to carry out at this point is identifying and studying competent job or role incumbents according to agreed-upon outcome criteria. The principal reason for this appears to be the difficulty of getting faculty agreement on what outcomes constitute competence or "success." In practice, this entire step is almost always overlooked, and in its place is substituted a priori conjecture about the qualities of competent individuals. One may value assumptions about the characteristics required for leadership, citizenship, creativity, and service and build a program of instruction to develop those qualities, but in the process one runs certain risks. If the assumptions are correct, the program may achieve its intended purpose, but if they are faulty, the training program may be inefficient, ineffective, or even detrimental. Some baseline of performance is needed against which to test assumptions or build theories of competence.

In many career settings, one can identify performance outcomes to identify top performers -- criteria such as accurate diagnosis for physicians, sales figures in industry, and peer and supervisor ratings of performance in managerial jobs. In the absence of these data, resource panels made up of job incumbents or other experts can usually identify the results or outcome criteria by which they judge a performer to be competent or outstanding. Educators, too, can be asked to supply a list of things they would look for to tell them whether alumni are effective or ineffective in meeting the institution's expectations for them. In the final analysis, however, it is usually easier to identify particular individuals whose performance we value than it is to identify agreed-upon outcome performance standards for the same persons. Since the goal is to identify competent performers, by whatever means, any of the above procedures is legitimate.

The thematic-analytic approach can be extended to include many sources of data. If one can gain access to samples of effective and less effective individuals, determined on the basis of agree-upon outcome criteria, they can be studied using a holistic framework, looking for individual differences in experience, perspective, and attainment that suggest notions of competency. Biographical retrospective techniques which yield detailed life histories (e.g., Dailey, 1971) are useful tools in identifying differences among individuals that suggest competencies. Differences in individual approaches to critical life situations can also be defined through detailed analysis of an individual's recollection of past events. Asking competent and less competent individuals to recall times when things went well and times when things went badly is an efficient way of obtaining information about abilities that yields important differences in performance. Analysis of retrospection is particularly applicable to education because it is tied to the individual, rather than the job, career, or special set of circumstances as the unit of analysis.

Carrying out any of these analytic processes requires a substantial instituional commitment, initial program support of both administration and faculty, and specialized research expertise. Because one or more of these factors is absent in most CBE programs, an adequate empirical determination of competencies is highly unlikely to occur. Nevertheless, some institutions are making the attempt at an empirical analysis of competent performers (e.g., the College for Human Services, Alverno College, and the University of Minnesota College of Pharmacy), involving job and task analyses and a comparison of successful with less successful performers. The data derived from the various attempted approaches tend to fall on a continuum between generic competencies and specific behavioral objectives, both of which may have been concurrently validated against job incumbents but have yet to be validated regarding the effects of a program of instruction. Nevertheless, at least these institutions are questioning whether the competencies they have derived bear any relationship to demonstrated performance.

Designing and implementing competency measures. The process of designing competency measures is a highly technical one that requires considerable time, resources, and patience for all concerned. Our survey of current undergraduate CBE programs has shown competency assessment to be a weak link in ensuring that program goals are met; and considering resource constraints, it is unlikely that this condition will be improved by creating massive institutional programs to develop competency assessment devices. The same problems that face employers in assessing applicant competencies (see Chapter 1) also face educators; for many reasons, guidelines for developing adequate measures are well established but are not followed in practice. As a result, CBE programs can probably benefit most from adopting existing testing technologies to their own unique situation rather than attempting to develop totally new assessment systems. The issue, then, becomes one of making informed choices among existing assessment alternatives.

The most important subset of assessment issues for educators are the four that form the framework for Section II of the present chapter: the nature of the competencies measured, the specificity of the assessment criteria, the amenability of the competencies to educational treatment and the validity of the competency measures. On one or more of these issues, current widely accepted testing practices should be examined by institutions to determine whether certain measures actually meet program needs. Of particular importance is the issue of validity, on which the claims of CBE for uniqueness largely rests.

At present, scores on standardized tests of academic achievement are commonly viewed as direct indicators of program effectiveness. Scores on the Scholastic Aptitude Test, for example, are taken as indicators of the effectiveness of secondary education, as the Graduate Record Examinations are viewed as indicators of how well an undergraduate experience has prepared students for professional school. This attitude persists in the presence of masses of evidence compiled by Price et al. (1974), McClelland (1973), Heath (1977), and others, citing a lack of relationship between traditional academic measures of aptitude, achievement, and intellectual ability, and a host of life and job performance outcomes. On the other hand, less often-considered measures of ego development (Loevinger & Wessler, 1970; Stewart, 1977) and moral reasoning (Kohlberg, 1980; Rests, 1976) have shown empirical and, in some cases, causal relationships between certain qualities of mind and of life success. These measures are also "competency based," in that they are referenced to life outcome criteria and in some cases have even been used as proxies for performance outcomes.

These measures, too, are all examples of generic competency tests; the qualities of mind which are assessed are related to a host of different outcomes that transcend specific situations, and in this sense they are criterion-referenced. Other measures of generic competency designed to capture performance dimensions relevant to some of the more commonly stated program goals (i.e., competencies such as "critical thinking" and "intellectual flexibility") are discussed in a later section of this chapter.

Another assumption which deserves investigation is whether the range of student responses to an assessment device is appropriate to the complexity of the competency being measured. No one would take seriously a multiple-choice test designed to assess public speaking skills, yet many people accept the idea that such paper-and-pencil tests can assess subtle intellectual process such as conceptual thinking. Multiple-choice, forced choice (e.g., "true-false"), and even short-answer essay tests may be appropriate measures of information recall or recognition, but they allow responses that are too limited for domonstrating other kinds of competency.

Respondent measures, as a class, also tend to have low criterion validity, which is to say that they are seldom related to important life and career outcomes (Ghiselli, 1966; Michel, 1968). By contrast, tests which may be classified as "operant," or free-response measures such as certain behavioral simulations and essay tests (given the caveat that the latter exmple is limited to written verbal response) provide opportunities for students to demonstrate competence in ways that more closely resemble later life performance requirements. Operant measures, other than those mentioned, include observer ratings in assessment center and classroom settings which, when properly conducted, show consistently higher predictive validity than respondent test techniques (MacKinnon, 1975).

Klemp (1979) has outlined some of the other important aspects related to choosing competency assessment instruments for use in educational settings. One of these is defining the domain of competencies to be assessed, including content validity and also what might be termed "process validity," or the extent to which the responses expected of students in assessment settings are isomorphic to the responses required in critical career or other life situations where a given competency is demanded. Another aspect is the various measurement properties of the instrument. This includes the sensitivity of the measure to variations in the degree to which a competency is possessed; the utility of the measure, or the amount of variation in response it permits; the uniqueness of the measure, or the lack of redundancy and overlap between measures of different competencies; reliability, or consistency of measurement; and validity, or the degree to which a test measures what it was designed to measure. Operant measures, though they are more time-consuming and difficult to score than respondent measures, tend to have greater sensitivity, utility, uniqueness, and validity than respondent measures and the better examples of operant measures have reliabilities well within the range of respectability. Thus, they are the choice to be preferred in assessing student outcomes in CBE programs.

Validating Measures of Program Effectiveness

There is not aspect of the seven-step competency-based assessment and program development process that a true CBE program can afford to ignore. Newly founded CBE programs should address each of the steps in accordance with their institutional missions. Existing CBE programs may find it extremely difficult to retrace their steps, but they should nevertheless begin to ask whether their programs of instruction bear any relationship to the life or career effectiveness of their graduates. The two critical steps are as follows: (1) developing measures of competency attainment which fulfill as many of the criteria for educational assessment as possible, and (2) evaluating curriculum, either concurrently or longitudinally.

We have stressed the integration of assessment and curriculum goals because, in the final analysis, no other kind of program evaluation is likely to establish a direct, useful link between measurable program outcomes and specific aspects of the program of instruction. Program evaluation is not something that should occur at the end of a student's career, nor should it be divorced from the ongoing process of student assessment in the classroom or the field. The same measures designed for use at the program level should comprise a significant part of the assessments used in program evaluation. Conversely, if generic measures normally used to assess student attainment just prior to graduation were made an integral part of curriculum design and ongoing evaluation (i.e., to assess change in the competencies over time) institutions would come closer to the ideal of developing a validated competency-based program of instruction.

A few programs have already been mentioned which have documented research on new measures which bear demonstrated relationships to a variety of later-life outcomes. Development or adaptation of measures by institutions for use in research programs funded specificially for the purpose of program evaluation is strongly encouraged. In particular, cooperative programs involving educational institutions and test developers are to be supported, provided that the research projects are geared to demonstrating the relationships between measurement, curriculum design and later performance.

Also to be encouraged is widespread dissemination of specific measures or measurement techniques for which empirical validity regarding adult outcomes has already been demonstrated. This dissemination should include more than just making the instruments available to faculty members or institutional assessors. It should also involve a strong component which provides explanations to students of what the measures are intended to measure, and how information from these measures can benefit them directly. In practice, most college testing programs regard students as "human subjects," providing feedback to them (when feedback is provided) only to the extent of reporting test scores, but not their relevance to competencies or performance outcomes. If assessment techniques are to be integrated into programs of instruction, their relevance needs to be explained to the students who are assessed as well as to faculty members who will either administer measures of competency attainment or make constructive use of the information provided by them.

Some New Measures of Educational Effects

Thus far, we have discussed a process for competency-based assessment and program development and some of its implication for assessment in CBE programs. In recent years, some measures of competencies related to liberal arts competency goals have been developed, not to evaluate individual students' performance, but rather to be used in research on the effects of different kinds of higher education. This research was begun by Winter, McClelland, and Stewart (1980) and by McBer and company, under a contract from FIPSE (Russell, 1979). Both groups of researchers had been concerned about the state of research and findings on the effects of higher education. Despite decades (or centuries) of luxuriant language about the goals of liberal education, there was little firm evidence about whether these goals were actually achieved. Winter and his colleagues believed that the evidence to data had been more a reflection of the tests used than of the efficacy of the liberal arts. For example, psychologists who study "thinking" or information-processing typically focus on relatively simple elements or processes. Winter and his colleagues felt that tests of liberal arts competencies should be based on what college students are actually required to do, rather than on what psychologists can easily (and cheaply) measure and score. For example, if one presumed competence of liberal education is forming and articulating complex concepts, then students should be given a mass of complex and confusing information and be assessed according to their ability to organize this information into their own coherent concepts. They should not be given examples of somebody else's concepts and asked to check which information applies to which concept, nor, worse still, be given examples of some artificial concept devised by a psychologist for laboratory research (e.g., combinations of shapes and colors, or even Chinese characters) and then asked to recognize whether further examples fit the concept.

In devising a test of complex concept formation, therefore, these authors tried to stay as close as possible to real-world tasks. They gave students two groups of brief, imaginative stories and then asked them to formulate and describe the differences between the two groups in any way they liked, a procedure called the Test of Thematic Analysis (Winter & McClelland, 1978). While the two groups of stories differed in many ways, the idea of the test does not depend on using these particular stories. Any two groups of complex information -- novels, paintings, systems of philosophy, or sets of scientific data -- could be substituted for the stories.

Winter and McClelland found that most students make rather vague and wandering comparisons, full of their own subjective reactions and emotions. It is as though they are not quite able to master the "blooming, buzzing confusion" of a complex world. Seniors at a prestigious, traditional liberal arts college, on the other hand, were able to make crisply-worded comparisons, using parallel structure, full of examples and qualifications, redefining the raw information where appropriate. They were able to recognize things that are functionally similar even though superficially quite different. In short, they are better able to bring the confusion and complexity of the stimulus data under cognitive control. Winter and McClelland believe this implies an ability to do the same in daily life. The differences between students as a whole and seniors hold up even after controling for differences in SAT scores and social class.

Another often cited goal of liberal education is to promote intellectual flexibility, i.e., the capacity to identify truth in all sides of a complicated and controversial issue. Instead of asking students whether they were intellectually flexible ("Always -- often -- occasionally -- seldom -- never: check one"), thus risking a response, Stewart (1976) developed a test called Analysis of Argument, which actually confronted students with an extreme and badly-argued point of view on an emotionally-laden issue (for example, an excerpt from a 1968 sermon by Norman Vincent Peale criticizing draft resistance, or from a 1978 speech by a New Hampshire politician blasting student critics of nuclear power plant construction). Students were told to attack the position. Senior at the liberal arts college made their attack in much more coherent, logical ways, as compared to freshman at their own college and students at other, more vocationally-oriented colleges. (Again, appropriate statistical controls were used to remove any effects of differences in intelligence and social class.) Without prior warning, students were also asked to defend the original argument. The usual reaction was to switch from a blanket condemnation to a blanket endorsement. The liberal arts seniors, however, were able to restrain themselves, to pick out a few elements of truth in the original statement while still being reasonably consistent with the logic of their previous attack.

The two preceding tests were designed with a major emphasis on achieving face validity: These tests do appear to measure what they claim to measure. Some other tests used in this line of research, however, were based on the familiar Thematic Apperception Tests (TAT) and may therefore have less face validity. Stewart devised a measure of *self-definition* (Stewart & Winter, 1974), which measures freedom from ascribed roles, an instrumental, effective style of thinking and acting, and a measure of *Stage of Adaptation to the Environment* (Stewart, 1977, 1979). McClelland combined TAT measures of high power motivation, low affiliation motivation, and high Activity Inhibition (or self-control) into a complex measure of the *Leadership Motivation Pattern* (McClelland, 1975).

All of the measures described by Winter et al. are generic in the sense that they measure liberal arts competencies across a wide range of different fields or subjects. Each is also supported with a detailed coding manual, and after brief self-training with the available practice materials, scorers can attain 85% or better agreement. Thus, these new measures are, both in principle and in practice, based on explicit criteria shared by assessors. Initital research studies also showed that these measures possessed educational validity. That is, they were designed to reflect changes associated with a liberal arts educational program, and in later validation studies, seniors at a liberal arts college showed scores on each measure that were signficantly higher than those of liberal arts freshman and higher than seniors at other kinds of colleges (a vocationally-oriented college and a community college). Research is currently being conducted both to replicate these findings with a longitudinal study of students at a liberal arts college, and also to pinpoint what institutional or educational features cause what changes, through a study of several quite different institutions of higher education (Russell, 1979).

Are these tests relevant to real-life adult performance, both in career roles and in the management of personal lives? There are two sources of data on the new measures. First, several of these measures have established (or growing) predictive validity credentials from previous research. For example, McClelland has shown that the leadership motive pattern predicts organizational ability and success among managers, particularly their capacity to generate high morale and performance among subordinates (McClelland, 1975; McClelland & Burnham, 1976). Self-definition predicts effective handling of stress and life crises (Stewart, 1978). In a recent study of Navy alcoholism counselors (Boyatzis & Burruss, 1977) scores on the Test of Thematic Analysis (modified to consist of two groups of case studies rather than two groups of imaginative stories) predicted success in dealing with clients, i.e., the proportion of alchololic clients who improved. It seems therefore, that the cognitive competency of complex concept formation plays an important role even in what is usually thought of as an interpersonal role that required empathy. Apparently, people have to create the right concepts in order to create the right concepts in order to help people effectively.

The second source of adult-life data is a longitudinal study of the class of 1964 graduates from a traditional liberal arts college (Winter, 1977; Winter, McClelland, & Stewart, 1980, Chapter 6). They had taken the TAT as freshmen (in September, 1960) and were followed up as adults ten years after college graduation, in 1974. (They did not take the Test of Thematic Analysis or theanalysis of Argument, because these had not yet been developed.) Over this fourteen-year span of life, the liberal arts competency measures based on even the freshman-year TAT showed important relationships to real-life adult performance and life satisfaction outcomes, holding constant other relevant background factors. For example, the Leadership Motive Pattern predicted being active as an officer in organizations; the Stage of Adaptation measure predicted career satisfaction, being an entrepreneur, and organizational membership--signs of mature adaptation to adulthood; and Self-definition predicted feeling energetic, having more income, and having a dual-career family.

The results are especially interesting if we compare the predictive power of these new TAT-based measures to that of two traditional measures of "competency": SAT scores and college grades (level of honors at graduation). Both of the traditional measures were actually negatively correlated with many aspects of later career success and satisfaction. The findings from this study confirm the results that Heath (1977) found at another traditional liberal arts college, and McClelland's general review of the poor power of SAT scores (and college grades) to predict to important real-life outcomes. Considering all the important life outcomes that were available for the researchers to measure, the new TAT-based liberal arts competency measures accounted for almost fifty percent more variation than did SAT scores and level of honors at graduation combined. Overall, the fourteen-year longitudinal correlations between test scores and life successes were significant, although not very high for any of the measures. This is natural, since crude measures of life outcomes within a heterogeneous group of young adults are affected by a variety of situational and historical factors, as well as by personality traits and educational competencies other than what was measured for the study. In this research, however, the real-life relevance of the new competency measures is substantially more impressive than is the relevance of SAT scores and college grades, two measures long assumed (and still assumed) to be related to postgraduate success and effectiveness.

VI. CONCLUSIONS

The scope of research for this chapter was an examination of CBE to determine how competencies have been defined and assessed and how well practice measures up in terms of employing state-of-the-art procedures and standards for assuring the accuracy of definitions and the validity of assessments. To this end, we chose to examine the strongest strain of the CBE movement--undergraduate and graduate programs, most of which were begun with special resources from FIPSE. Due in large measure to FIPSE's efforts to bring people with a sophisticated understanding of issues of competency definition and measurement together with new arrivals on the CBE scene, the level of understanding of the state of the art is generally higher among these practitioners. At least one institution in the set, Alverno College, is attempting to move beyond content validation to construct and criterion referenced validation of its measures; and several other programs have become explicit about both the outcomes they expected and the criteria to be used in evaluating whether those outcomes had been achieved. While competency-based teacher education (CBTE) had been around longer, it had been appraised over a decade of growth, the conclusion being that CBTE programs were not built upon an improved understanding of teacher competence. Moreover, CBTE assessments were at best only face valid or content valid in relation to taxonomies of teaching tasks or behaviors, the educational effectiveness of which was suspect. Through a review of literature, personal contact and formal interviews, we also assessed what was occurring in two other emerging areas of CBE: general education at the high school level and vocational and technical education. These checks led us to conclude that in selecting undergraduate and graduate programs, we had indeed selected the strongest strain of CBE for close examination. Our conclusions and recommendations, however, apply to CBE programs in general.

Looking back at FIPSE's original goals, agency staff wished to fund "outcome-oriented," "student-centered" programs. As presented in this chapter, the strongest test of whether these goals were met would be provided by educational programs whose competency goals bore a demonstrated relationship to adult performance and which succeeded in developing competencies in their students that were specified in the program goals. In principle, FIPSE's objectives would have been satisfied if institutions merely took seriously the goals for learning printed in their catalogues and paid attention to the progress of all students, not just "the best and the brightest," "the average student," or "the traditional student."

In looking at CBE's potential, this might be said to be a weaker test of success: Were the student outcomes at graduation that which educators had intended and attempted to measure? Assuring student competence, individually or collectively, in terms of an institution's goals (as opposed to course objectives of individual professors) was considered a progressive leap forward were it to occur, or a way of recapturing a lost horizon of higher education. The test of success of outcome-oriented CBE programs, in this case, stops at the exit gate: Did the program succeed in meeting its own goals for academic competence? What precisely was accomplished? What did students learn? Or, in the jargon of program evaluators, what was the "value added" by the experience?

The implication of the word "program" is a unified vision of what the institution wishes to accomplish for its students, and a set of overarching goals that must be evaluated. Program objectives, then, are greater than the sum total of individual course objectives set by faculty. Program evaluation means more than course evaluation and requires examining whether the institutional goals have been achieved. In terms of this weaker test of success of CBE programs, what do we know about program outcomes? We found that very few programs in our sample assessed progress of students toward institutional goals; most of the programs assessed only at the level of individual courses. Curricula may, in fact, be coherent and efficacious in terms of meeting institutional goals, but most programs surveyed have not been evaluated to determine whether or not that is the case.

For those CBE programs which claim to be educating students for extra-institutional competence, verification of claims requires that such programs ultimately be submitted to the stronger test of CBE. Decades of effort to develop sound procedures for verification have yielded sound methodologies for validating assessments, noted here and elsewhere (e.g., Chapters 1 and 3). Evaluating the efficacy of pedagogical approaches in promoting competence in the workplace hinges upon the adequacy of assessment procedures for capturing critical dimensions of job competence.

Given that checks have been made to assure that assessments: (a) measure what they set out to measure (face, content, and construct validity); (b) measure progress on what was actually taught (instructional validity); and (c) successfully sort the competent from the incompetent (concurrent validity), the best procedure at our disposal for verifying that we have taught and assessed something critical to competent performance in extra-institutional settings is predictive validation. Do students who do well on the institution's assessments perform competently as workers, citizens, parents, friends, or whatever the roles for which educational programs were intended as preparatory? The test must fit the claim.

To date, neither the weaker nor the stronger claim of CBE's effectiveness in promoting competence has been put to the test. The prospects for the future, moreover, are dim unless program evaluation that encompasses predictive validation is undertaken. The logic of predictive validity is simple, but the execution is technologically complex. It requires an investment in time before results that may affect the program are known. CBE programs must first have reached the full stage of implementation and have graduates who occupy target roles.

As a consequence, practically speaking, concurrent validation becomes the requirement in practice. But before an institution can even begin to evaluate its program along those lines, its goals and assessment criteria must be specific and explicit. Few programs have reached the necessary level of explicitness and specificity of criteria, so few programs are currently in a position to conduct concurrent or predictive validation studies. Moreover, few programs even aspire to doing so. In the absence of concurrent or predictive validation, institutional claims to the effect that their graduates, having been assessed for competence in tasks deemed important by educators, will perform competently (or more competently than others) in later-life roles, cannot be verified.

There exist a number of barriers to conducting the research necessary for CBE to reach its full potential. First and foremost is the lack of understanding of the importance of this kind of research among funders and among most CBE practitioners. FIPSE, the agency with the best understanding of the issues, cannot sponsor research. NIE, with major federal level responsibility for educational research, has a very small division with interest in the topic and few staff able to adequately appraise proposals in the area. The fact that predictive validity studies take years to complete also imposes the requirement of long-termed sustained interest on the part of funding agencies. Most research foci at the federal level, especially in education, change frequently as leadership comes and goes, decreasing the likelihood of the necessary sustained commitment.

These barriers to launching a program of systematic research aimed at improving educational assessment and pedagogy through CBE are formidable, but not insurmountable. Where programs are ripe for this kind of evaluation (be they in mainstream education or the world of work) and where there is willingness on the part of the institution to undertake it, agencies need to cooperate to assure that funding is provided. Such research presents a rare opportunity to answer perennial questions of educators and employers, and for that matter, the public: What enables competent performance, and how should competency be assessed?

A second barrier to research is the fact that many emerging CBE programs lack people who have the interest or technical capacity to do an adequate job of defining competence. Curriculum development is the first concern, assessment of attainment is the second, and validation of competency is a distant last-place finish. Accordingly, few CBE programs retain specialists with a grasp of the methodological issues required for defining which competencies are important and how they may be assessed. The world of work doesn't provide many good answers as to which competencies are important to obtain prior to employment (See Chapter 1), so educators are left with most of this responsibility. In Section V we described some approaches to defining competencies that have been pursued in a variety of educational settings. Where such research is not feasible, educators must nevertheless make some determination as to what the competencies might be. Judgments can be based upon tradition, philosophies and theories of education, or personal experience, but all must be put to an empirical test at some later point, as several CBE programs and some traditional programs are doing. This requires an up-front investment in personnel with the capability and willingness to execute a plan of research to verify the importance of the competencies stressed in the program goals.

A third barrier, perhaps the most pernicious of all, is the seeming logic of a priori theory and face validity. Traditional programs and many CBE programs see no need to validate their goals and measures, because it is "obvious" that what they are doing through their curricula is creating educated men and women, better able to tackle life's uncertainties and the challenges of the workplace. Such understandings based on tradition and casual experience were responsible for the conclusion of our earlier age that the earth was flat. Without the persistence of bold scientists and explorers against cries of blasphemy and threat of pain, the modern sciences of astronomy and geography would not have been possible. Today, educators are being forced into examining cherished assumptions in the face of mounting evidence of a lack of relationship between important life outcomes and grades or other traditional measures of educational attainment. The whole notion of obviousness should be abandoned. Without at least a cursory glance at the workplace, we cannot assume that a program, no matter how thoughtfully it is conceived, provides a student adequate preparation for work; the program must be evaluated against student outcomes. The potency of face validity applies, in particular, to competency assessment. The fact that a program uses a variety of innovative and seemingly "real-life" approaches to testing often leads observers to conclude that the tests are relevant to later-life performance. But in most cases, _ad hoc_ paper-and-pencil tests, simulations and other behavioral demonstrations simply do not measure what they seem to be measuring and need to be examined.*

Finally, a fourth major barrier to implementing CBE is the costs involved in transforming a traditional program into one that is competency-based. In theory, the design and execution of a competency-based program along the lines of the program development process described in Section V of this chapter should not be that much more costly than developing a traditional program. The greater costs have come from attempting to retrofit traditional programs to competency-based standards. In the long run, it is much cheaper to start a competency-based program from scratch, where there are fewer philosophical and procedural barriers to overcome, than it is to build a CBE program from a traditional base. The innovative competency-based programs being developed by the University of Minnesota, the Navy and the American Management Association testify to the fact that outcome-oriented education programs, begun with an empirical research base, are feasible. We also observe that it is not realistic to expect that traditional programs converted into competency-based ones will validate competency goals without a fundamental re-examination process. The two undergraduate programs which are closest to demonstrating the later-life relevance of competencies--Alverno College and the College for Human Services--had to redesign curriculum and assessment substantially, and not without struggle. The message here is that CBE cannot fulfill its promise if it is treated as old wine in new bottles. Even the vineyard must be planted anew.

---

*See, especially, Chapter 3 on the assessment center in educational settings.

In terms of educational policy, the findings in this chapter have implications for many areas of concern. Particularly regarding the relationship between education and work, the issue of paramount importance for this chapter, there are three themes that pull together our thoughts about how CBE may provide capable and productive members of society.

1. *Providing conceptual clarity at the federal level.* We have found that the majority of CBE programs are characterized by the adoption of the language of competency, including "competency-based" measures of attainment and redefined program objectives. Compared with traditional education, some of these changes are rhetorical, while others have to do with the process of measurement. As far as can be determined, however, the majority of CBE programs in higher education and almost all other

CBE programs are no more "outcome-oriented" in their desire to demonstrate the relationships between competencies and effectiveness in the workplace than their traditional counterparts. Much of this behavior, moreover, is shaped by federal funding sources: little emphasis is placed on outcomes research, and program implementation is more clearly rewarded than program validation. But federal agencies can change funding contingencies and thereby reshape institutional behavior. This process would be facilitated by a renewed emphasis on program evaluation, particularly around evaluating the ultimate worth of a program in contributing to national interests. Before putting massive funding into new educational programs, agencies should sponsor analyses of what skills are needed--not "armchair" analyses, but empirical analyses of competency requirements in the workplace. This research should be stimulated through requests for proposals open to academic institutions as well as institutions outside the academic community who may be closer to the occupations of interest.

Research, too, must continue on competency measures which predict outcomes of interest in postacademic life. There are currently few tests in widespread use that have proven useful in measuring competencies having a demonstrated relationship to life and work outcomes, and more such measures need to be made available. Though there are many different approaches to developing useful measures which measure important generic skills, such as the empirically-derived Test of Thematic Analysis, are probably the most efficient means of assessing broad areas of student attainment. Here, funding agencies can help by demanding that assessment of outcomes be an integral part of program development in CBE, thereby providing an incentive to use state-of-the-art assessment techniques.

Policymakers must also begin to make distinctions between what we have termed minimal and optimal competence. Minimal competencies, including most specific knowledge and skills, are fairly easy to define and measure, and there is a danger of these competencies overtaking curricula. Vocational and technical programs are especially susceptible to this effect. The problem with this is that employers are ordinarily equiped to provide training for job-specific knowledge and skill, within reasonable bounds of cost, but are not well positioned to promote more generic, optimal competencies also required for competent job performance. Lack of adequate attention on the part of the education system to promoting optimal competencies cannot be remediated in work environments. A job poorly done by the schools in this regard leaves students with little recourse throughout their careers to overcome the deficiencies. For developing most optimal competencies, long periods of time and concentrated effort are required. As Dewey noted in Democracy and Education (1916), the workplace, with its deadlines and efficiency requirements, does not allow for postponing tasks while one goes about following up questions which arise in the course of their execution. There is precious little opportunity for the exploration and intensive investigation needed to develop optimal competence. Employers see generic intellectual and interpersonal skills as comprising the greatest gap between their needs and what the schools are producing. Therefore, prudential concern for national productivity suggests the need for schools to focus on generic abilities more directly in curriculum development rather than to assume, without examination, that because they were included as goals, they were achieved.

Needless to say, the problem of minimal specific knowledge and skills crowding out time in schools for the development of critical optimal competencies looms as a danger of the back-to-basics movement and minimum competency testing for high school graduation. Tests of specific knowledge and skills such as balancing checkbooks and reading bus schedules can drive the curriculum if that is all that is measured for accountability purposes. CBE in higher education, too, must become aware of the ramifications of choosing to emphasize survival skills over excellence. In all cases, care must be taken to ensure that in attempts to lay a floor, the ceiling doesn't collapse.

2. Encouraging sophistication in the field. Concern for improving the quality of instruction at all levels of the educational system has led to considerable investments in program evaluation. Education leads all fields in terms of energies expended on this kind of research. One of the lessons learned thus far is that the absence of evaluation activity or poor quality of evaluations at the state and local level is not due

to stubbornness, but rather to a lack of technical capacity. Successful efforts to change the situation have been a product of demonstrating the power of evaluation for program improvement and organized efforts to upgrade the research skills of state and district agencies.

We recommend that part of the capacity-building agenda be improving the level of understanding in the field of issues of validity and how validation studies can be put to work to improve the quality of instruction. Improved capacity to deal with these issues is first needed at the federal level to support field initiatives. Then, one of the roles that NIE, for example, could assume is providing technical assistance to practitioners. By encouraging funded programs to seek out people with the skills needed to carry out proper validation studies, federal agencies would raise the awareness of educators around issues of program evaluation, measurement validation, and efficient curriculum design. Likewise, the dissemination of materials on competency-based program design, test construction, and the fundamentals of concurrent and longitudinal program evaluation would be a valuable service. Lastly, providing support, within educational institutions themselves, for evaluation units charged with testing claims of program relevance to career outcomes would provide concrete feedback that would enable faculty to reassess goals and educational processes.

An additional means of providing CBE programs with sophistication around what outcomes are important to assess would come from cooperative research studies on jobs and careers carried out jointly by educational institutions and employers. Particularly for schools whose graduates are largely employed by local business and government, such studies would satisfy two needs at once. They would make educators aware of the competencies actually required for a variety of work roles, and they would help employers gain a better grasp of employment selection criteria. Presumably, students who have been certified competent in certain critical skill areas should be acceptable to employers requiring the same mix of skills. If this could be accomplished, it would be a most efficient advance in educational practice.

3. Rethinking the relationship between education and work. The concept of competence implies the ability to adapt to change. If knowledge and skill requirements of jobs should become obsolete, or the number of jobs available in particular fields should shrink, what aspects of competence enable the individual to acquire new knowledge and skills? This question alone leads us to conclude that if the role of formal education is to enable individuals to cope with change, the competencies that should be developed by schools should be generic rather than specific. Moreover, as noted earlier, employers provide on-the-job learning opportunities through which many specific competencies are acquired through experience in performing tasks. Accordingly, we need to be more open to the idea that education does not occur only in schools. If this seems obvious, our speaking of "education" and "work" as two different worlds reveals acceptance of a false dichotomy.

There is a need to reconsider the division of responsibility for education and training between schools and work settings. If specific, job-related competencies are more efficiently developed through on-the-job training programs, why not subsidize employers to provide this type of educational experience? This is currently being done through internship and work-study programs in many fields of endeavor, and has the potential of enhancing access to permanent jobs. All agencies with interests in manpower training and development should take stock of their investments in light of our findings. As communication between schools and the world of work increases, we need to begin to assess what areas of competence are implicated by the dialogue and then decide where educational dollars can best be spent.

In asking these questions, we cannot neglect a concern for equity. At present there is no mesh between concerns for equal employment and equal educational opportunities. The schools are the domain of focus in thinking about equal educational opportunity and the work place is the domain for examining equal employment opportunity. Assuring that minority populations gain access to the work setting seems to be where current policies begin and end. The concept of equal employment opportunity also encompasses promotional opportunities or opportunities to grow and develop on the job. Populations entering the world of work without generic competencies important for optimal performance and advancement are severely handicapped. There is a need, then, to search for ways of diagnosing such deficiencies among workers and a need to discover efficient ways of promoting generic competencies for these populations that do not require the forfeiture of the beachhead they have gained in the world of work.

In conclusion, traditional education and training programs have seldom suffered from an inability to help students or workers acquire specific knowledge and skills. Furthermore, home, community and work settings are also rich in opportunities for their acquisition. Returns on investments in CBE are therefore quite small if the focus is on specific knowledge and skills. Huge amounts of time, effort and money would have been expended for very small gains. On the other hand, if CBE programs could demonstrate superior capability in promoting generic competencies critical to improving national productivity and quality of life, returns on investments would be incalcuable.

REFERENCES

Alamprese, J. Joint dissemination review panel submission. Submitted to the National Institute of Education by Syracuse Research Corp., 1979.

Alverno College Faculty. Assessment at Alverno College. Milwaukee: Alverno College, 1979.

Alverno College Faculty. Liberal learning at Alverno College. Milwaukee: Alverno College, 1976.

Alverno College Office of Evaluation: Careering after college: Establishing the validity of abilities learned in college for later success. Progress report submitted to the National Institute of Education (project #NIE-G-77-0058), January 1979.

Borich, G. D. Sources of invalidilty in measuring classroom behavior. Instructional Science, 1977, 6(3), 283-318.

Chickering, A., et al. Competency-based: Alverno College. In Developing the College Curriculum, Washington, DC: Council for the Advancement of Small Colleges, 1977.

Clasby, M., et al. Laws, tests and schooling: Changing contexts for educational decision-making. Educational Policy Research Center, Syracuse Research Corp. Prepared under Contract #OEC-0-72-5017 for the U.S. Office of Education, October 1973.

Cohen, A. C. The citizen as the integrating agent: Productivity in the human services. HEW Human Services Monograph Series, No. 9, September 1978.

Cort, H. R., & Sammons, S. L. Diagnosis and evaluation of lawyering competencies in law school clinics. Washington, DC: Antioch School of Law, 1978.

Cross, K. P. Accent on learning: Improving instruction and reshaping the curriculum. San Francisco: Jossey-Bass, 1976.

Dailey, C., & McClelland, D. C. Professional competencies of human service workers. Boston: McBer and Company, 1974.

Dewey, J. *Democracy and education*. Toronto, Ontario: The Free Press, 1966.

Feldman, K. A., & Newcomb, T. M. *The impact of college on students*. San Francisco: Jossey-Bass, 1973.

Flanagan, J. The critical incident technique. *Psychological Bulletin*, 1954, *51*(4), 327-358.

Friedman, M., & Mentkowski, M. *Instrument validation in a competence-based college curriculum: An empirical illustration*. Paper presented at the American Education Research Association meetings, April 1980.

Gamson, Z., College IV, Grand Valley State Colleges: Competence-based education in a modular curriculum. In G. Grant (ed.), *On Competence*. San Francisco: Jossey-Bass, 1979.

Goldstein, I. *Training: Program development and evaluation*. Monterey, CA: Brooks/Cole, 1974.

Grant, G., et al. *On Competence*. San Francisco: Jossey-Bass, 1979.

Grossman, J. The College for Human Services. *American Education*, May 1979, *15*(4), 11-16.

Heath, D. H. Academic predictors of adult maturity and competencies. *Journal of Higher Education*, 1977, *48*(6), 613-32.

Higdon, H. *The crime of the century: The Lobe and Leopold Case*. New York: Putnam's, 1975.

Huff, S. M. Report of the Special Legal Issues Conference of the Multi-State Consortium on Performance-Based Teacher Education. In B. Levitov (ed.), *Licensing and Accreditation in Education: The Law and State Interest*. Lincoln, NE: Nebraska Curriculum Development Center, 1976.

Irby, A. J.. Statewide assessment of basic skills. *American Association for Higher Education Bulletin*; 1979, No. 5, 11-16.

Jones, N. S. The Antioch School of Law: A case study. In G. Grant (ed.), *On Competence*, Vol. II. Syracuse, NY: Syracuse Reserach Corp., 1978.

Kane, J., & Lawler, E. Methods of peer assessment. *Psychological Bulletin*, 1979.

Knowles, M. *The modern practice of adult education: Androgogy versus pedagogy*. New York: Association Press, 1970.

Mager, R. *Preparing instructional objectives*. CA: Fearson Publishers, Inc., 1975.

McClelland, D. C. Testing for competence rather than for "intelligence." *American Psychologist*, 1973, *28*(1).

McClelland, D. C., & Winter, D. G. *Motivating economic achievement*. New York: The Free Press, 1969.

McDonald, F. J. *Executive summary report, beginning teacher evaluation study, phase II*. Princeton, NJ: Educational Testing Service, 1976.

Medley, D. M., et al. *Assessment and research in teacher education: Focus on PBTE*. Washington, DC: American Association of Colleges for Teacher Education, 1975.

Mentkowski, M., et al. *Developing a professional competence model for nursing education*. Paper presented at the American Education Research Association meetings, April 1980.

Merwin. *Performance-based teacher education: Some measurements and decision-making considerations*. Washington, DC: American Association of Colleges for Teacher Education, 1973.

Messick, S. The standard problem: Meaning and values in measurements and evaluation. *American Psychologist*, 1975, *30*, 955-966.

Moses, J. L., & Byham, W. C. (eds.) *Applying the assessment center method*. New York: Pergamon Press, 1977.

Nickse, R. S. *Assessing and certifying adult life skills: The New York State External High School Diploma Program*. Belmont, CA: Pearson-Pitman, 1980 (in press).

Nortcutt, N., et al. *Adult functional competency: A summary*. Austin, TX: University of Texas at Austin, 1975.

Oleson, V. Overcoming crises in a new nursing program: Mt. Hood Community College. In G. Grant (ed.), *On Competence*. San Francisco: Jossey-Bass, 1979.

Report on the Core Curriculum. Cambridge, MA: Harvard University, Office of the Dean, 1978.

Ryan, K., et al. An interview with Ralph Tyler. *Phi Delta Kappan*, March 1977, 544-547.

Schmieder, A. A. Profile of the states in competency-based education. *PTBE* Multi-State Consortium on Performance-Based Teacher Education, November 1974, *3*(5).

Shavelson, R., & Atwood-Russo, N. Generalizability of measures of teacher effectiveness. *Educational Research*, *19*(3), June 1977, 171-183.

Shavelson, R., & Dempsey, N. Generalizability of measures of teaching behavior. *Review of Educational Research*, Fall 1976, *46*(4), 553-611.

Silber, D. L., et al. The SIU medical curriculum: Systemwide objectives-based instruction. *Journal of Medical Education*, June 1978, *54*(6), 473-479.

Stewart, A. Analysis of argument, in McClelland, D. C. *The impact of college on personal characteristics and life outcomes*, Cambridge, MA: Harvard University, 1976.

Stewart, A. J., & Winter, D. G. Self-definition and social definition in women. *Journal of Personality*, 1974, *42*, 238-259.

Tyler, R. Personal communication. 1980.

White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, *66*, 297-333.

Winne, P. H., & Marx, R. W. Reconceptualizing research on teaching. *Journal of Educational Psychology*, 1977, *69*(6), 668-678.

Winter, D. G. *An introduction to LMET theory and research*. Report prepared for the U.S. Department of the Navy. Boston: McBer and Company, 1979.